

From: Li, Ruixiang
Sent: Tuesday, October 30, 2001 2:41 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application NO: 09/775,181

Please do a standard search of SEQ ID NOS: 1, 2, and 4 against both the commercial and interference nucleic acid databases.

Thank you very much!

Ruixiang Li
GAU 1646
CM1 8E18
Mail Box 10C01
306-0282

*Edward Hart
Technical Info Specialist
STIC / Biotech
CM1 12C14 Tel: 305-9203*



SEARCH REQUEST FORM

Scientific and Technical Information Center

Access DB#

53994

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: _____	NA Sequence (#) <u>1</u>	STN <u>Reverse to NA</u>
Searcher Phone #: _____	AA Sequence (#) <u>2</u>	Diag <u>Reverse to NA</u>
Searcher Location: _____	Structure (#) _____	Questel Orbit _____
Date Searcher Present: <u>10/30/01</u>	Bibliographic _____	Dr. Luma _____
Date Timed out: <u>11/5/01</u>	Litigation _____	Lexis Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>04</u>
Client Prep Time: _____	Patent Family _____	WWW Internet _____
File Time: _____	Other _____	Other Specialty _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 06:44:17.; Search time 3196.64 Seconds
(without alignments)
17651.768 Million cell updates/sec

Title: US-09-775-181-1
Perfect score: 3648
Sequence: 1 atgggagcattgcttacc.....gggatatgttttaaaagttag 3648

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_em:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_bal:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vi:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_vl2:*
- 60: gb_htg1:*
- 61: gb_htg2:*
- 62: gb_htg3:*
- 63: gb_htg4:*
- 64: gb_htg5:*
- 65: gb_htg6:*
- 66: gb_htg7:*
- 67: gb_htg8:*
- 68: gb_htg9:*
- 69: gb_htg10:*
- 70: gb_htg11:*
- 71: gb_htg12:*
- 72: gb_htg13:*
- 73: gb_htg14:*
- 74: gb_htg15:*
- 75: gb_htg16:*
- 76: gb_htg17:*
- 77: gb_htg18:*
- 78: gb_htg19:*
- 79: gb_htg20:*
- 80: gb_htg21:*
- 81: gb_htg22:*
- 82: gb_htg23:*
- 83: gb_htg24:*
- 84: gb_htg25:*
- 85: gb_pri:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_rol:*
- 95: gb_rol2:*
- 96: gb_in4:*
- 97: gb_pri10:*
- 98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match			DB ID	Description
		Match	Length	%		
1	1634.4	44.8	4595	85	AB032962	Homo sapi
2	1304.4	41.2	151218	79	AL355587	AL355587 Homo sapi
c 3	1443	39.6	185146	79	AL355542	AL355542 Homo sapi
4	1077.8	29.5	1857	85	AB052146	AB052146 Macaca fa
5	903	24.8	11117	80	AL358932	AL358932 Homo sapi
6	903	24.8	19526	79	AL139821	AL139821 Homo sapi
c 7	227.2	6.2	143878	80	AL359974	AL359974 Homo sapi
c 8	225.6	6.2	159596	79	AL161654	AL161654 Homo sapi

Db 601 GTGGTGGCCGTTGGTGTGCAAGTGCAGAACGCTCAACCTCAGCTCAGAGAAGAAAC 660
Qy 2673 tgggacccacgaacatcgatgttagaagatctctcagtgctcatagcaagcccaagga 2732
Db 661 TGGCACCCACGACATCGATGTTACAGAACTCTCAGTGTCTATAGCAAGCCCAAGGA 720
Qy 2733 gaagactcttgattagctgggaaaccccaacagcagggtgtggaagaacgcactaaac 2792
Db 721 GAAGACTCTTGGATTAGCTGGGAAACCCCAACAGCAGGTTGTGGAAGACGCATTAATC 780
Qy 2793 ccagaaacctttccaaaagataaagagacaaacagaaatcactcaaatcttgataaac 2852
Db 781 CCAGAAACCTTTCCAAAAGATAAAGACAAACAGAAATCACTCAAAATTTCTATAACAC 840
Qy 2853 agagactaaagatctctgcccacaaactcaaatctctgcagagagcgaagaagcctca 2912
Db 841 AGAGACTAAAGATCTCTGCCCCCAAACTCAATCTTCGCGAGGAGCCAGAAAGCCTCA 900
2913 gaaatctgggattatgaacacaaaggggtcaacccccaccactgccaatctgacctgaa 2972
901 GAAATCTGGATTATGAACAACAAGGGTCAACCCACCCTGCCAATTTGACCTGAA 960
Qy 2973 ccaggcaccacccagatgaagacaactttgacattggggaggtgtgctctgggaggt 3032
Db 961 CCAGGCACCACCAGATGAAGACAACTTTGACATTTGGGGAGTGTGCTTTGGGAGGT 1020
Qy 3033 ttatgacctgacctgctgctgcttcagatcaatacaaaagtccaagcagctatctat 3092
Db 1021 TTATGACCTGACCCCTGCTGCTGCTTCAGATCAAAAGTTCAAAAGCAGCTATCTAT 1080
Qy 3093 tgtgctctgaaatggagaaaaacccccacttttcttaaaaggagaaatctcaccacaa 3152
Db 1081 TGTGGCTTCTGAAATGAGAAAAACCCCACTTTTCTTAAGGAGAAATCTCACCACAA 1140
Qy 3153 gctaagcagctgaggtttgtcagcaatccaatcagaagcgcataagatgaagctgaagt 3212
Db 1141 GCCTAAGGCAGCTGAGGTTTGTGAGCAATCCAAATCAGAGCGCATAGATAAGGCTGAAGT 1200
Qy 3213 atgctctggagagccaaagccagctcatttggaaatgagaaagctttgtattccaa 3272
Db 1201 ATGCTTTGGAGAGCCAAAGCCAGTCCATTTTGGAGATGAGAAAGCTTTTGATTTCCAA 1260
Qy 3273 gactcagctctcccagagagggcaaaagaggaggaacggaggttcagcctcgtgcagccaa 3332
Db 1261 GACTCCAGTTCTCCAGAGAGGGCAAAAGAGAGAGAGAGAGGAGTCCGCTGTCAGCCAA 1320
3333 tgtgtgctggcagagcgaagaactgcccccccaagctgtagcatcaaaaacagagaa 3392
1321 TGTGTGCTGGCAGAGCCGAAGAACTGCCCCCAAGCTGTAGCATCAAAACAGAGAA 1380
Qy 3393 tgaatactcaacaaataggacaccagagaaaaaagacatctctctgaggagaatgt 3452
Db 1381 TGAATACTCAACCAATAGCACACCAGGAAAAAAGACATCTTCTCTCAGGAGATGT 1440
Qy 3453 gctggtcctctaataactcaagtaataacttcagacacccctttaacatcacgagaggt 3512
Db 1441 GCGTGGCTCTATAAATCAAGTAATAACTTCCAGCAACCTTTTAACATCACGAGCAGAGGT 1500
Qy 3513 ttatcctggaggttgagaccagcagctcaaccaatgctggaagaagttagctttacc 3572
Db 1501 TTGTCTCTGGAGTTTGAGACCCAGCTCAACCAAAATGCTGGGAAGAAGTGTAGCTTTACC 1560
3573 tgcctctctgcttaagtccaatagatagcagggtcccttagaagaagaagatctggga 3632
Db 1561 TGCTCTCTCTAGTCAATAGATAGTAGCAGGCGCTTAGGAAAGAGAGGTCTCGGA 1620
Qy 3633 tagtttaaagtgtag 3648
Db 1621 TAGTTTTAAAGTGAG 1636

RESULT 2

AL355587 151218 bp DNA HTG 15-APR-2001
LOCUS Homo sapiens chromosome 10 clone RP11-561H23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 9 unordered pieces.
AL355587
ACCESSION AL355587
VERSION AL355587.8 GI:13660940
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton, J.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13624988.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA561H23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147891 bases at least Q40
Consensus quality: 149178 bases at least Q30
Consensus quality: 149893 bases at least Q20
Insert size: 150418; sum-of-contigs
Quality coverage: 4.69x in Q20 bases; sum-of-contigs Quality
coverage: 5.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 21608: contig of 21608 bp in length
* 21609 21708: gap of 100 bp
* 21709 48919: contig of 27211 bp in length
* 48920 49019: gap of 100 bp
* 49020 57539: contig of 8520 bp in length
* 57540 57639: gap of 100 bp
* 57640 69961: contig of 12322 bp in length
* 69962 70061: gap of 100 bp
* 70062 101299: contig of 31238 bp in length
* 101300 101399: gap of 100 bp
* 101400 114233: contig of 12834 bp in length
* 114234 114333: gap of 100 bp
* 114334 126584: contig of 12251 bp in length
* 126585 126684: gap of 100 bp
* 126685 138216: contig of 11532 bp in length
* 138217 138316: gap of 100 bp
* 138317 151218: contig of 12902 bp in length.
Location/Qualifiers
1. .151218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-561H23"
/clone_lib="RPC1-11.2"
1. .21608
/note="assembly_fragment:00445
fragment_chain:1
clone_end:T7
misc_feature
1. .21608
/note="assembly_fragment:00445
fragment_chain:1
clone_end:T7
FEATURES
Source
1. .151218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-561H23"
/clone_lib="RPC1-11.2"
1. .21608
/note="assembly_fragment:00445
fragment_chain:1
clone_end:T7

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AL355542.9 GI:13624986
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185146)
Plumb,B.
Direct Submission
Submitted (12-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 14, 2001 this sequence version replaced gi:13624506.

COMMENT

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA257J14
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 175310 bases at least Q40
Consensus quality: 178563 bases at least Q30
Consensus quality: 180883 bases at least Q20
Insert size: 183346; sum-of-contigs
Insert size: 145596; 18.0% error; agarose-fp
Quality coverage: 4.36x in Q20 bases; sum-of-contigs Quality
coverage: 7.64x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 11771: contig of 11771 bp in length
* 11772 11871: gap of 100 bp
* 11872 61631: contig of 49760 bp in length
* 61632 61731: gap of 100 bp
* 61732 79327: contig of 17596 bp in length
* 79328 79427: gap of 100 bp
* 79428 83309: contig of 3882 bp in length
* 83310 83409: gap of 100 bp
* 83410 109528: contig of 26119 bp in length
* 109529 109628: gap of 100 bp
* 109629 112808: contig of 3180 bp in length
* 112809 112908: gap of 100 bp
* 112909 118905: contig of 5997 bp in length
* 118906 119005: gap of 100 bp
* 119006 133743: contig of 14738 bp in length
* 133744 133843: gap of 100 bp
* 133844 143213: contig of 9370 bp in length
* 143214 143313: gap of 100 bp
* 143314 158185: contig of 14872 bp in length
* 158186 158285: gap of 100 bp
* 158286 162061: contig of 3776 bp in length
* 162062 162161: gap of 100 bp
* 162162 164747: contig of 2586 bp in length
* 164748 164847: gap of 100 bp
* 164848 167658: contig of 2811 bp in length
* 167659 167758: gap of 100 bp
* 167759 170009: contig of 2251 bp in length
* 170010 170109: gap of 100 bp
* 170110 173787: contig of 3678 bp in length
* 173788 173887: gap of 100 bp
* 173888 176350: contig of 2463 bp in length
* 176351 176450: gap of 100 bp
* 176451 179441: contig of 2991 bp in length

* 179442 179541: gap of 100 bp
* 179542 182272: contig of 2731 bp in length
* 182273 182372: gap of 100 bp
* 182373 185146: contig of 2774 bp in length.

FEATURES

source
1..185146 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPC1-11.1"
1..11771
/note="assembly_fragment:00155
fragment_chain:1"
11872..61631
/note="assembly_fragment:00988
fragment_chain:1"
61732..79327
/note="assembly_fragment:01331
fragment_chain:1"
79428..83309
/note="assembly_fragment:03102
fragment_chain:1"
83410..109528
/note="assembly_fragment:02919
fragment_chain:1"
109629..112808
/note="assembly_fragment:01757
fragment_chain:1"
112909..118905
/note="assembly_fragment:01162
fragment_chain:1"
119006..133743
/note="assembly_fragment:01246
fragment_chain:1"
133844..143213
/note="assembly_fragment:01266
fragment_chain:1"
143314..158185
/note="assembly_fragment:03087
fragment_chain:1"
158286..162061
/note="assembly_fragment:00095"
162162..164747
/note="assembly_fragment:00400"
164848..167658
/note="assembly_fragment:00509"
167759..170009
/note="assembly_fragment:01477"
170110..173787
/note="assembly_fragment:01903"
173888..176350
/note="assembly_fragment:02102"
176451..179441
/note="assembly_fragment:02396"
179542..182272
/note="assembly_fragment:02514"
182373..185146
/note="assembly_fragment:03322"

BASE COUNT 50876 a 38769 c 39206 g 5480 t 1815 others
ORIGIN

Query Match 39.6%; Score 1443; DB 79; Length 185146;
Query Local Similarity 99.1%; Pred. No. 0;
Matches 1493; Conservative 0; Mismatches 10; Indels 4; Gaps 4;

Qy 2143 cgggacgagctgaaaaaactctatgcccaactggaatataataaagaagaatgac 2202
Db 164377 CAGGACGAGCTGAAAAAATCTATGCCCACTGGAATATATAAAGAAACAGATGATC 164418
Qy 2203 acaaacacccccaccctccagaaaaaacggtgctcgaagaaggccctaggtcgttcacatc 2262
|||||


```
/clone="Qcce-20671"
/sex="male"
/tissue_type="cerebellum cortex"
/clone_lib="macaque brain cDNA library Qcce"
/dev_stage="adult"
213..1166
CDS
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB19001.1"
/db_xref="GI:11611573"
/translation="MLQKSLSVIASAKERTLGLAGTGTAGVEESAKSOKPLPKDKET
NRHNSDNTETKDPQNSNPAEPKPKQSGIMKQORVNPPTANSDLNPGATOMKD
NFDIGVCWEVYDLTPGPVPSKVKHVSIVASEMKNPTLSLKEKSHRKAKEV
CQSNQKCTDKASVENCWQGLQIGQSLIDEKHFISKTPLVQERKEENGQPHAAKYCAG
QSELPFPKAVASVENLNQIGHQEKKTSFENVRGYSNNNSNFOPLMSRAEYCPW
EFTTPQPNAGRVALPVSALSASIAGRPKREVWDTFKV"
BASE COUNT      629 a 395 c 414 g 419 t
ORIGIN

Query Match      29.5%; Score 1077.8; DB:85; Length 1837;
Best Local Similarity 95.7%; Pred. No. 7.6e-242;
Matches 1119; Conservative 0; Mismatches 47; Indels 3; Gaps 1;

QY 2480 aaacggaagtcacgtacgtacccacagaaagccaaagaggagagacacagaaatt 2539
Db 1 AAACGGGAAGTCCAGTAGGCTACCCACAGAAAGCCAAAGAGGAGGAGGTGACAGAAATT 60
QY 2540 ccacactggaatccctgtcgggttaaaactaacacaaaactaaagaagacagcgagg 2599
Db 61 CCACACTGGAATCCTGTCGAGTAAGAACTGACACAAAACCTAAAGAGACACTGAGG 120
QY 2600 ctgagtcacaggagtcggtgctgtgtgtaagtcagcaagcgctcaaacctcagct 2659
Db 121 CTGAGTCCACGGAGTCGCTGCGTGTGTGCAAGTCAGCAAGTGCTCACAACTCAGCT 180
QY 2660 cagagaagaatactggcaccacagacacatgatgttacgaagctctcagtgctatag 2719
Db 181 CAGAGAAGAGCCCTGGACACCCACGACGTCATGTTTACAGAAGTCTCTCAGTGTCATAG 240
QY 2720 caagcgcaaggagaagactcttgattagctgggaaaccccaacagcagcggtgtgaag 2779
Db 241 CAAGTGCCAGGAGAACACTCTTGATATTAGTTGGGAAACCCCAACAGCGCGTGGAAAG 300
QY 2780 aacgcactaaatccacagaacactctgcccagaagataagagacaaacagaaatcacctaa 2839
Db 301 AAAGTGCTAAATCCAGAACCTCTGCCAAAGATAAAGAGACAAACAGAAATCACTCAA 360
2840 attctgatacacagagactaaagatcctgcccccaaaactcaaatcctcgaggagagc 2899
Db 361 ATTCTGATAACAGAGACTAAAGATCCTACCCCCCAAACTCAAAATCCTCGGAGGAGC 420
QY 2900 caagaaagcctcagaaatctgggattatgaacacaaaggggtcaacccaccactgcca 2959
Db 421 CAAGAAAGCCTCAGAAATCTGGGATTATGAACAAACAAAGGGTCAACCCACCCTGCCA 480
QY 2960 attctgacctgaacccaggcaccacccagatgaaggacaaactttgacattggggagtggt 3019
Db 481 ATTCTGACCTGAACCCAGGCGCCACCCAGATGAAGGACAACTTTGACATTGGGGAGGTGT 540
QY 3020 gtccttgggaggttatgacctgacctgacctgctgtgctgtccttcagaatacaaaagtccaa 3079
Db 541 GTCCCTGGGAGGTTTATGACTGACCTGACCCCTGCTGCTGCTTTCAGATCAAAAGTTCAA 600
QY 3080 agcagctatctattgtgctctgaaatggagaaaaacccacttttctcttaagagaga 3139
Db 601 AGCATGATATCTATTGTGGCTTCTCAATATGGAGAAAAACCCCACTCTTTCCTTAAAGGAGA 660
QY 3140 aatctaccacaagccttaagcagctgaggtttgtcagcaatccaatcagaagcgcatag 3199
Db 661 AATCTCACCACAGCCCTAAGGCAGCTGAAGTTTGTGAGCAATCCCAATCAGAAGTGCACAG 720
QY 3200 ataaggtgaagtatgcctttgggagagccaaagcccgctccattttggaagatgagaagc 3259
```

```
Db 721 ATAAGCGTGAAGTATGCTTTGGGGGATCCAAAGCCCAAGTCCATTTTGAAGATGAGAAGC 780
QY 3260 ttctgattccaagactccagttctccagagagggcacaagaggaacacgaggtcagc 3319
Db 781 ATTTCATTTTCAAGACTCCAGTCTCTCCAAAGAGAGGGCAAAAGAGAGAGCAACGAGGTGAGC 840
QY 3320 ctctgtcagccaatgtgtgtcgtgggcagagcgaagaactgcccccaaaagctgtagcat 3379
Db 841 CTCATGCAGCAAAAGTGTGTGCTGGGCAGAGTGAAGAACTGCCCCCAAAAGCTGTAGCAT 900
QY 3380 caaaaacagagaatgaaaaatctcaaccataagacacccagggaaaaaaagacatctctt 3439
Db 901 CAAAAACAGAGAAATGAAAAATCTCAACCAAAATAGACACACCAGGAAAAAAGACA---TCCT 957
QY 3440 ctgaggagaatgtgcgtggctcctataactcaagtcaacttcacgtccagcaactttaacat 3499
Db 958 TTGAGGAGAAATGCGTGGCTCTCTATAACTCAAGTAACTTCCACCAACCTTTAATGT 1017
QY 3500 cacgagcagaggtttgtccttgggagtttgagaccccgctcaaccaaagtctggaagaa 3559
Db 1018 CACGAGCAGAGGTGTGCTTGGAGTTTGAGACCCCGAGCTCAACCAAAATGCTGGAAGAA 1077
QY 3560 gtgtagctttacctgcctcttctgtcttaagtgcacaaataagatagcagggccttaggaag 3619
Db 1078 GTGTAGTCTTTACCTGTTTCTTCTCTTAAGTCAAGTAAAGATAGCAGGGCCTCGGAAG 1137
QY 3620 aagagatctgggagtagttttaaagtgtag 3648
Db 1138 AGAAGTCTGGGATACTTTTAAAGTGTAG 1166
```

```
RESULT 5
AL358932
LOCUS      AL358932      111117 bp      DNA      HTG      23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION  AL358932
VERSION     AL358932.3
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sims,S.
Direct Submission
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926667.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA395P8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 90261 bases at least Q40
Consensus quality: 99226 bases at least Q30
Consensus quality: 104490 bases at least Q20
Insert size: 108717; sum-of-contigs
Insert size: 178772; 24.0% error; agarose-fp
Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality
coverage: 1.42x in Q20 bases; agarose-fp
-----
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known, and their order in this sequence record is

Db 97278 AAGGGAAGCCGACCCAGCAGCGGGTGGAGCCCTCTGCTTCGACATCTCTGGCTCC 97337
QY 181 tggagcgcgtccaccgatggcaccatctggcgcagaaactgcgcagagaggtgcccattg 240
Db 97338 TGGAGCGGCTCCACCGATGGCACCATTCTGGCGCAGAAACTGCCGAGAGGTGCCCAATG 97397
QY 241 gacgtggcctcttacctctacacccgggggactccaccagctgaagcagagccaaactgctcc 300
Db 97398 GAGCTGGCCTCTTACCTCTACACCGGGGACTCCACACAGCTGAAGGAGCCAACTGCTCC 97457
QY 301 gcccgtacagttggcgggctgcggggaagtggcagccctggcagcgcgcacccc 360
Db 97458 GCGCGCTACAGTTGGCGGCTTCGCGGGAAGTGGCCAGCCCTGGCCAGCGCACCC 97517
QY 361 tccttcacccggcgtggacacactgacacgcgcacacacacttctcaactgagctg 420
Db 97518 TCCTTGCACCGGGGCTGGACACTGACACAGCCACCACTTCTCAACTGATGCTG 97577
QY 421 cagagcaataagtgcgggagcagaacttgcaggagcactgattgggtaccagggcgtg 480
Db 97578 CAGAGCAATAAGTCCGGGAGCAGAACTTGCAGAGCAGCTGGATTGGTACCAGCGCTG 97637
QY 481 gigtgagcctctggaggcagccagcatctccggggcgccatcacccttcagcacc 540
Db 97638 GTCGTGAGCCTTCTGGAGGGGAGGCCAGCATCTCCGGGGGCCATCATCCTTCAGCAC 97697
QY 541 gattcgtgtccgcacccgcccacaggtcttctccagccacgcgcgcagagagcgcgc 600
Db 97698 GATTCGCTCTCCGACCGGCCCCACAGGTCTTCTCCAGGCCACGCGCGAGAGCGCG 97757
QY 601 atctgtctcagaactgtctctctccgcacccacccacttggccaaagcactctggagacc 660
Db 97758 ATCTGTCTCAAGACTGTCTCTCCGACCCACCTTGGCCAGCCACTTGGCAGGAC 97817
QY 661 gattggttccagcgcctccggcgaagtggagggccacacttaccacgcgcgcgcgcgc 720
Db 97818 GAGTGTCTCCAGCGGCTCCGGCGCAAGTGGAGGGCCCCACTTACCGCGCGGCCCAAT 97877
QY 721 caggggcccgccggcgtggccacagctggcgcgcgaagcagcgcgcgcgcgcgcgcgc 780
Db 97878 CAGGGCGCCCGGGCTGGGCCACAGCTGGCGGCGCAAGGAGCGGCTCGCGGGGACAG 97937
QY 781 agccacttaagtgtctcgccttatctgtgagtcgagagcagcgcgcgcgcgcgcgcgc 840
Db 97938 AGCCACTTCAAGTGTCTCGCCTTATCTGAGTGGAGAGCGGAGTACAAAGCCCGG 97997
QY 841 tgctggttactcttctctccatctacggttgcagcctaacctgctgcgcgcgcgcgcgc 900
Db 97998 TGGCTGGTTACTCTTCTCTCCATCTACGGGTTGCAGCCTTAACCTGGTCCCGGAATTC 98057
901 agg 903
Db 98058 AGG 98060
RESULT 7
AL359974/c
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP13-236A4, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION AL359974
VERSION AL359974.4 GI:9864577
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
i (bases 1 to 143878)
Sims,S.
Direct Submission
TITLE
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT:

On Aug 22, 2000 this sequence version replaced gi:9795088.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: B236A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 130224 bases at least Q40
Consensus quality: 135875 bases at least Q30
Consensus quality: 138890 bases at least Q20
Insert size: 14478; sum-of-contigs
Insert size: 205930; 23.5% error; agarose-fp
Quality coverage: 3.17x in Q20 bases; sum-of-contigs Quality
coverage: 2.50x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved

* 1 12130: contig of 12130 bp in length
* 12131 12230: gap of 100 bp
* 12231 12274: contig of 8044 bp in length
* 12275 20374: gap of 100 bp
* 20375 23062: contig of 2688 bp in length
* 23063 23162: gap of 100 bp
* 23163 26227: contig of 3065 bp in length
* 26228 26327: gap of 100 bp
* 26328 33260: contig of 6933 bp in length
* 33261 33360: gap of 100 bp
* 33361 35571: contig of 2211 bp in length
* 35572 35671: gap of 100 bp
* 35672 38834: contig of 3163 bp in length
* 38835 39334: gap of 100 bp
* 39335 44371: contig of 5437 bp in length
* 44372 44471: gap of 100 bp
* 44472 47729: contig of 3258 bp in length
* 47730 47829: gap of 100 bp
* 47830 52657: contig of 4828 bp in length
* 52658 52757: gap of 100 bp
* 52758 54776: contig of 2019 bp in length
* 54777 54876: gap of 100 bp
* 54877 57327: contig of 2451 bp in length
* 57328 57427: gap of 100 bp
* 57428 60599: contig of 3172 bp in length
* 60600 60699: gap of 100 bp
* 60700 64861: contig of 4162 bp in length
* 64862 64961: gap of 100 bp
* 64962 71930: contig of 6969 bp in length
* 71931 72030: gap of 100 bp
* 72031 75423: contig of 3399 bp in length
* 75430 75529: gap of 100 bp
* 75530 78541: contig of 3012 bp in length
* 78542 78641: gap of 100 bp
* 78642 80961: contig of 2320 bp in length
* 80962 81061: gap of 100 bp
* 81062 83093: contig of 2038 bp in length
* 83100 83199: gap of 100 bp
* 83200 95484: contig of 12285 bp in length
* 95485 95584: gap of 100 bp
* 95585 97735: contig of 2151 bp in length
* 97736 97835: gap of 100 bp
* 97836 102251: contig of 4416 bp in length
* 102252 102351: gap of 100 bp
* 102352 131392: contig of 29041 bp in length

```
* 131393 131492: gap of 100 bp
* 131493 137581: contig of 6089 bp in length
* 137582 137681: gap of 100 bp
* 137682 143878: contig of 6197 bp in length.
FEATURES
    source
        1..143878
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="RP13-236A4"
            /clone_lib="RPCI-13.1"
            1..12130
                /note="assembly_fragment:01047"
                fragment_chain:1
                12231..20274
                    /note="assembly_fragment:00007"
                    fragment_chain:1
                20375..23062
                    /note="assembly_fragment:01237"
                    fragment_chain:1
                23163..26227
                    /note="assembly_fragment:01103"
                    fragment_chain:1
                26328..33260
                    /note="assembly_fragment:00266"
                    fragment_chain:1
                33361..33571
                    /note="assembly_fragment:01006"
                    fragment_chain:1
                35672..38834
                    /note="assembly_fragment:00201"
                    fragment_chain:2
                38935..44371
                    /note="assembly_fragment:01057"
                    fragment_chain:2
                44472..47729
                    /note="assembly_fragment:00179"
                    fragment_chain:2
                47830..52657
                    /note="assembly_fragment:00444"
                    fragment_chain:3
                52758..54776
                    /note="assembly_fragment:00414"
                    fragment_chain:3
                54877..57327
                    /note="assembly_fragment:00586"
                    fragment_chain:4
                57428..60599
                    /note="assembly_fragment:00641"
                    fragment_chain:4
                60700..64861
                    /note="assembly_fragment:00717"
                    fragment_chain:5
                64962..71930
                    /note="assembly_fragment:01116"
                    fragment_chain:5
                72031..75429
                    /note="assembly_fragment:00156"
                    fragment_chain:5
                75530..78541
                    /note="assembly_fragment:00233"
                    fragment_chain:5
                78642..80961
                    /note="assembly_fragment:00273"
                    fragment_chain:5
                81062..83099
                    /note="assembly_fragment:00279"
                    fragment_chain:5
                83200..95484
                    /note="assembly_fragment:00515"
                    fragment_chain:5
                95585..97735
                    /note="assembly_fragment:00654"
                    fragment_chain:5
                97836..102251
                    /note="assembly_fragment:00675"
                    fragment_chain:5
                102352..131392
                    /note="assembly_fragment:00719"
                    fragment_chain:5
                131493..137581
```

```
misc_feature      /note="assembly_fragment:00767"
137682..143878
/note="assembly_fragment:00852
clone_end:SP6
vector_side:right"
BASE COUNT      44271 a 25988 c 25788 g 45414 t 2417 others
ORIGIN

Query Match      6.2%  Score 227.2; DB 80; Length 143878;
Best Local Similarity 93.0%; Pred. No. 1.2e-42;
Matches 238; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1112 gaaggggtccggtacagcatatttcaggaagtcacaaagatgtgtcagaagaagcctatg 1171
|||||
Db 89285 GAAGGGGTCCGGATCAGCATATTTTCAGGAAGTACAAAGAGTGTGTCTCAGAGAAGCCTATG 89226

Qy 1172 tctgcctaccttcagagagggtgcccttctgtctgatgacagccatgcttcgtcc 1231
|||||
Db 89225 TCTGCCTACCTTGAGGAGGGCTGCCCTTCTGTGCTGTGATGACAGCCCATGCTTCGTCC 89166

Qy 1232 aggaagataagtatttcagcttgcctcatctcttcacagccttcacaaagcagcctcgg 1291
|||||
Db 89165 AGGAAGATAAGTATTAGCACTTGCCATCATCTCTTCCAAAGGCTGTGTATGTGCTCG 89106

Qy 1292 acttcgttagcatgtgtgtctaccacttttcgaaagcagcctcggcggcatcgg 1351
|||||
Db 89105 ACTTCGTTAGCATGCTGTGTGCTTACCACCTTTCGAAAGCAAGTAAACCCAGGAACCC 89046

Qy 1352 gccttacctgttggg 1367
|||||
Db 89045 TGGTATGATCCTGTA 89030

RESULT 8
LOCUS      AL161654 159596 bp DNA HTG 11-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-59G22, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION  AL161654
VERSION     AL161654.8 GI:13620309
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 159596)
AUTHORS     Burton,J
TITLE       Direct Submission
JOURNAL
COMMENT      Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
            On Apr 12, 2001 this sequence version replaced gi:13567947.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: BA59G22
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 157954 bases at least Q40
            Consensus quality: 158428 bases at least Q30
            Consensus quality: 158729 bases at least Q20
            Insert size: 158896; sum-of-contigs
            Insert size: 159333; 6.0% error; agarose-fp
            Quality coverage: 6.69x in Q20 bases; sum-of-contigs Quality
            coverage: 6.67x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
```

* consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 24276: contig of 24276 bp in length
 * 24277 24376: gap of 100 bp
 * 24377 31732: contig of 7356 bp in length
 * 31733 31832: gap of 100 bp
 * 31833 67342: contig of 35510 bp in length
 * 67343 67442: gap of 100 bp
 * 67443 70588: contig of 3146 bp in length
 * 70589 70688: gap of 100 bp
 * 70689 81295: contig of 10607 bp in length
 * 81296 81395: gap of 100 bp
 * 81396 111898: contig of 30503 bp in length
 * 111899 111998: gap of 100 bp
 * 111999 126181: contig of 14183 bp in length
 * 126182 126281: gap of 100 bp
 * 126282 159596: contig of 33315 bp in length.

FEATURES

source

1. .159596
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone_lib="RPCI-11.1"
 1. .24276

misc_feature
 /note="assembly_fragment:01962
 fragment_chain:1
 clone_end:77
 vector_side:left"

misc_feature
 24377..31732
 /note="assembly_fragment:01615
 fragment_chain:1"

misc_feature
 31833..67342
 /note="assembly_fragment:02112
 fragment_chain:1"

misc_feature
 67443..70588
 /note="assembly_fragment:00781
 fragment_chain:1"

misc_feature
 70689..81295
 /note="assembly_fragment:00892
 fragment_chain:1"

misc_feature
 81396..111898
 /note="assembly_fragment:02522
 fragment_chain:1"

misc_feature
 111999..126181
 /note="assembly_fragment:01840
 fragment_chain:1"

misc_feature
 126282..159596
 /note="assembly_fragment:01914
 fragment_chain:1"

clone_end:SP6
 vector_side:right"

BASE COUNT 54172 a 29596 c 28660 g 46068 t 700 others
 ORIGIN

Query Match 6.2%; Score 225.6; DB 79; Length 159596;
 Best Local Similarity 92.6%; Pred. No. 2.7e-42;
 Matches 237; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1112 gaagggtccgacagcatatttcagggaagacaaaagatgtgcagaagagcctatg 1171
 |||

Db 102607 GAAGGGTCCGATCAGCATATTTGAGGAAGTGTGCAAGAGCCTATG 102548

QY 1172 tctgctacctgcaggagggtccctctgtgctgatcacagccatcttctgctc 1231
 |||

Db 102547 TCTGCTACTTGCAGGAGGGCTGCCCTTCTGTGCTGATGACAGCCATCTCGTCC 102488

QY 1232 aggaagaatagatttacgacttgccatctctccaaaggcctgtgtatgctctcg 1291
 |||
 Db 102487 AGCAAGTAAGTATTTACGACTTGCATCTCCCTCCAAAGCCCTGTGTATCTCTCG 102428
 |||
 QY 1292 acttcgttagcatgctgggtgtctaccactttgcacaaagacagcattccgggcatcg 1351
 |||
 Db 102427 ACTTCGTTAGCATGCTGGTGGTCTACCACTTTGCAAGCAAGGTAACCCAGGAACCC 102368
 |||
 QY 1352 gccttatcctgttggga 1367
 |||
 Db 102367 TGGTTATGATCTCTGTA 102352
 |||

RESULT 9

AL354976

LOCUS

AL354976

DEFINITION

Homo sapiens chromosome 10 clone RP11-2016, *** SEQUENCING IN

PROGRESS ***; 4 unordered pieces.

ACCESSION

AL354976

VERSION

AL354976.7 GI:13561182

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 186964)

AUTHORS

Plumb,B.

TITLE

Direct Submission

JOURNAL

Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Apr 8, 2001 this sequence version replaced gi:12330777.

COMMENT

----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA2016
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 185320 bases at least Q40
 Consensus quality: 186184 bases at least Q30
 Consensus quality: 186460 bases at least Q20
 Insert size: 186664; sum-of-contigs
 Inset size: 169440; 6.9% error; agarose-fp
 Quality coverage: 7.22x in Q20 bases; sum-of-contigs Quality
 coverage: 8.22x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 149917: contig of 149917 bp in length
 * 149918 150017: gap of 100 bp
 * 150018 153565: contig of 3548 bp in length
 * 153566 153665: gap of 100 bp
 * 153666 162836: contig of 9171 bp in length
 * 162837 162936: gap of 100 bp
 * 162937 186964: contig of 24028 bp in length.
 Location/Qualifiers
 1. .186964
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone_lib="RPCI-11.1"

```
misc_feature 1..149917
/note="assembly_fragment:00866.0"
misc_feature 150018..153565
/note="assembly_fragment:01186"
misc_feature 153666..162836
/note="assembly_fragment:01998"
misc_feature 162937..186964
/note="assembly_fragment:00147
vector_side:right"
vector_end:SP6

BASE COUNT 64194 a 34055 c 33553 g 54858 t 304 others
ORIGIN

Query Match 3.1%; Score 112; DB 79; Length 186964;
Best Local Similarity 95.8%; Pred. No. 1.1e-15;
Matches 115; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1002 ttccagtgatgcgaattaaagccctaggattcttcttgagccctatgagtcatttg 1061
|||||
179738 TTCTATGTTGATGCCAATTAAAGGCCCTAGGATTCGTTCTTGGAGCCCTATGAGTCATTTG 179797
|||||

Qy 1062 caaagcaggattctatcctcgtgagcttaccagtgaaacaacttccgagaaaggggtcc 1121
|||||
Db 179798 CAAGCAGGATTCATCATCTCGAGCTTACCACTGAACAACCTTTCGGAGTAAGTGCCC 179857
|||||

RESULT 10
AL445244/c
LOCUS Homo sapiens chromosome 10 clone RP13-99J21, *** SEQUENCING IN
DEFINITION PROGRESS ***, 6 unordered pieces.
ACCESSION AL445244
VERSION AL445244.4 GI:10800726
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sims,S.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 13, 2000 this sequence version replaced gi:10798428.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bB99J21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 158004 bases at least Q40
Consensus quality: 159033 bases at least Q30
Consensus quality: 159555 bases at least Q20
Insert size: 159890; sum-of-contigs
Insert size: 144185; 13.1% error; agarose-fp
Quality coverage: 5.59x in Q20 bases; sum-of-contigs Quality
coverage: 6.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 43907: contig of 43907 bp in length
* 43908 44007: gap of 100 bp
* 44008 55206: contig of 11199 bp in length
* 55207 55306: gap of 100 bp
* 55307 67735: contig of 12429 bp in length
* 67736 67835: gap of 100 bp
* 67836 73636: contig of 5801 bp in length
* 73637 73736: gap of 100 bp
* 73737 100153: contig of 26417 bp in length
* 100154 100253: gap of 100 bp
* 100254 160390: contig of 60137 bp in length.
FEATURES
Location/Qualifiers
source 1..160390
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-99J21"
/clone_lib="RPCI-13.1"
misc_feature 1..43907
/note="assembly_fragment:01136
clone_end:T7
vector_side:left"
misc_feature 44008..55206
/note="assembly_fragment:01384
fragment_chain:1"
misc_feature 55307..67735
/note="assembly_fragment:01898
fragment_chain:1"
misc_feature 67836..73636
/note="assembly_fragment:01997
fragment_chain:1"
misc_feature 73737..100153
/note="assembly_fragment:01657
fragment_chain:1"
misc_feature 100254..160390
/note="assembly_fragment:01063
fragment_chain:1
clone_end:SP6
vector_side:right"
BASE COUNT 55625 a 29765 c 28364 g 45935 t 501 others
ORIGIN

Query Match 3.0%; Score 108; DB 81; Length 160390;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 aggggtgctgaagtgacataatcttcagaaagtggaattgacaaatgctcaagt 960
|||||
Db 136464 AGGGGTGTGATGAAGTTGACATAAATCTTCAGAAAGTGGACATTCGCTCAAGT 136405
|||||

Qy 961 gatgctggttttcagggaactcataaagccacctcaacaattcagag 1008
|||||
Db 136406 GATGCTGTTTTCAGGAACCTCATAAATGCCCTCAACAATTCAGAG 136357
|||||

RESULT 11
I66494/c
LOCUS I66494 7218 bp DNA PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
source 1..7218
/organism="unknown"
```


numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E29 lies between and overlaps with cosmids E63 and E39 on the AseI-E genomic restriction fragment.

```

FEATURES             Location/Qualifiers
     source            1..26477
                        /organism="Streptomyces coelicolor A3(2)"
                        /strain="A3(2)"
                        /db_xref="taxon:100226"
                        /clone="cosmid E29"
     gene              1..6356
                        /gene="SCE63.01"
     misc_feature      complement(1..101)
                        /note="Overlap with Streptomyces coelicolor cosmid StE63."
     CDS                <1..6356
                        /gene="SCE63.01"
                        /note="SCE63.01, partial CDS, cdaps3, CDA peptide
                        synthetase III, len: >332aa; Constitutes the the majority
                        of cdaps3, CDA peptide synthetase III, part of the
                        calcium-dependent antibiotic (CDA) biosynthetic cluster
                        from Streptomyces coelicolor. CDA is a peptide antibiotic
                        which is synthesised non-ribosomally by a putative
                        multifunctional peptide synthetase enzyme. This partial
                        CDS encodes a subunit of this enzyme suspected to be
                        responsible for the addition of 2 amino acids to the
                        peptide antibiotic. This ORF overlaps the upstream
                        (cdapsII) by one base indicating possible translational
                        coupling. Contains two Pfam matches to entry PF003501
                        AMP-binding, AMP-binding enzyme, a Pfam match to entry
                        PF00668 DUF4, Domain of unknown function, a Pfam match to
                        entry PF00975 Thioesterase, Thioesterase domain and two
                        Pfam matches to entry PF00550 pp-binding,
                        Phosphotantetheine attachment site. Each PF00501 contains
                        a Prosite match PS00455 Putative AMP-binding domain
                        signature. Each PF00550 contains a Prosite match PS00012
                        Phosphotantetheine attachment site."
                        /codon_start=3
                        /transl_table=11
                        /label=cdaps3
                        /product="CDA peptide synthetase III"
                        /protein_id="CAB38876.1"
                        /db_xref="GI:4490979"
                        /translation="RURPAETLTGFLRRLQEQEARLIDHOWGLAEIQRWAGSGELFDP
                        TMVFNFFNLNSGRDPPGAADADLTVLGVRSKDQMHYPGLGLALPRETLRFLSLG
                        LPQVDFNRYAIAAFRVALRVLDAPRTVGAALLDPEVRGTVLEKSGSDVVRP
                        AERFTDLFEQVARTPGKALIPADGRLTVAELDAANRLARLIVELGVCYPERHVAVA
                        CGRLTFLVGMALVLAGNAYVPVDEYPPDRIHMIQDDADPALVLTSDVDRIEGEE
                        VCGRLTFLVMDPDNTGSLGRSGTALTDADRAAPLIVGHPAYVIYTSGTGRPKGVVV
                        EHRLSAFYHRCRSGQAPDISGLVSMQASFSQVSGSLHAPILISGCVKRLTDRLRA
                        ETAGSGEPFRHATFMKGTPLSLALLATMPPEVAPSGTTLTGGEELRGEILAPNREAAQ

```

DVTVMVYGPTEATGHCLEHWIAPDRTVEPQGVPIGTGTHPEGVRYVYVLDLSALRPVAPGQ
DGEYVLQVQJLYGRGLTAEFTADPCAGPSRMYRTGDVHANNEGELVFFAG
ADROYKLGRVYELBGEIEAAGVGGPQVAAVILREDRPQDQRLVAVVPPQGHWEA
AARARALSLPDMFMPSPAFVLDLGLSLPNKGLDRAALPAPTYTGTAGRAPRTPAEE
ILCDYALVLSLPGVTVDDPFDLGLSLTAAQRLMLHLLEGPSPTNITPTVLRISGP
ALAVVLAGAGRAVYDPRFLPGLSTLAQRULVADGVPVRFETAEADYEAAD
LPDLAARAAFDLVVGHSHSLRTTFTEDEBGRQVVPADGVPVRFETAEADYEAAD
LARAARAFDLGAIEYIPRARLJLSEBHEVILLVHHITASDAMSGPLAQDLTAAYTA
CRGADGAPAMPTPVQVQYDALWQOEILGDDTDPTLAGRLQYAWKQQLAGLPLERLDLP
TDRPAPADHTGDGVEEAFPALDHLTRTELARATDTLYVWLOQAALTLTRHIGAGE
DILGIPVAGRTDDADTHLVGFENVTLVLRDTSNPTFRLLTRVRDQDLTAYTHQD
LPERLVEALNPTRSILHPLLFQVWLSURSTAPRADCEGAPALPGLRVSGTGGAAA
TAAKQVLDGVSVTERAAADHTPDGAVGLDFTDLDRGTAGQDLVRLVLAADAAHP
DPLSLTIDLVLPRRHRVVEWNATAKGLATATLPelferrvrerpgeaevavcdpsh
SYAELNANRANRLVARGAGPERLVALPRLSAELPFAVLAVAKAGAAYLPDPAFL
PAERIGCTDDAAVPAVLTAAVAGLPTDQVPRLLIDDEBPAAGGDAADLTADRAL
APLLPAGHVAAYVSCTTGTRGPQVTVHSGLPALLDIFTSQLDVVPGSRVHLHLSNAP
DGGFWELAMGLNGAALVVVEGTVPGFALAAALVARHRVTHAATPAVLQILPEIGALP
AGTTVLVAATCEPELVVWASGRALMNSYGTPTVTCATMSAPLAGAAVPIIGRPTA
DTAGVYLDALQPVVPGPEGLYVRGGLYVGRPLSLYGRPSLTAGREVACPGFAGCVMYR
TGDLVHRHADGGLVDRDQTVOKLGRNRPVPAETAAVYAGLPGVQAAVLVREDTPVR
DRRLVGVVDPDAGSLGCDALQALRQALSLPEVMPVAPVALLDLPALTNGKLDHRLP
APEYTVGRSPRTPREALCRLEFVYLGLELVGLDDGDFDGLGSHLLAILRVRVRA
WEEGELVRDLFAAPTIYDALVLAARAGREPMERLLPRAAGTAPRVFCVHGPSGMS
XCSGLVIRHLLPQPIYVQQAQLDGDPGLQLOEMAAEYADIVRQTPGEPYRLILG
WSLGNVAFAMERANRAGQEVVELLAFLDAPVPRACAGPEAPLAEVAHNLRLDAGFDV
AEELGTGPREPTARYRAFLNAGADPMGLRDEAEALAAVLEVMNNAALMRGHTPCTYDGL
DLVLAARADGDKLARAGESWRPHVRGRITRERGVDADHLGLVQSDNALAVIGRALA
GRDPAITGNASAAVPEITGGVTAAMPSPPEPAPSPESLDSTEV^a
567. .1794
/gene="SCB63.01"
/note="Pfam match to entry PF00501 AMP-binding,
AMP-binding enzyme, score 385.20, E-value 6.6e-112"
957. .992
/gene="SCB63.01"
/note="PS000455 Putative AMP-binding domain signature."
2046. .2238
/gene="SCB63.01"
/note="SCB63.01"
/note="Pfam match to entry PF00550 pp-binding,
Phosphopantetheine attachment site, score 78.40, E-value
4.6e-20"
2115. .2162
/gene="SCB63.01"
/note="PS000112 Phosphopantetheine attachment site."
2283. .3480
/gene="SCB63.01"
/note="Pfam match to entry PF00668 DUF4, Domain of unknown
function, score 335.30, E-value 6.9e-97"
3789. .4980
/gene="SCB63.01"
/note="Pfam match to entry PF00501 AMP-binding,
AMP-binding enzyme, score 448.70, E-value 5.2e-131"
4167. .4202
/gene="SCB63.01"
/note="PS000455 Putative AMP-binding domain signature."
5232. .5424
/gene="SCB63.01"
/note="Pfam match to entry PF00550 pp-binding,
Phosphopantetheine attachment site, score 76.90, E-value
1.2e-19"
5301. .5348
/gene="SCB63.01"
/note="PS000112 Phosphopantetheine attachment site."
5490. .6219
/gene="SCB63.01"
/note="Pfam match to entry PF00975 Thioesterase,
Thioesterase domain, score 139.00, E-value 8.5e-38"
6360. .7178
/gene="SCB29.0"
/note="SCB29.0"
/gene="SCB29.0"
/note="SCB29.0, probable hydrolase, len: 272aa; similar
to many eq. TR:052809 (EMBL:AJ223998) from the vancovycin
to many eq. TR:052809 (EMBL:AJ223998) from the vancovycin

biosynthesis cluster of *Amvcolatopsis orientalis* (276 aa)
 fasta scores; opt: 89%, z-score: 1028.9, E(): 0, (50.9%
 identity in 271 aa overlap) and TR:O67982 (EMBL:AF003947)
 pcal; single polypeptide combining 3-oxoadipate
 enol-lactone hydrolyzing and 4-carboxymuconolactone
 decarboxylating activity from *Rhodococcus opacus* (400 aa)
 fasta scores; opt: 288, z-score: 335.4, E(): 2.3e-11,
 (32.0% identity in 236 aa overlap). Contains Pfam match to
 entry PF00561 abhydrolase, alpha/beta hydrolase fold,
 score 77.80, E-value 2.2e-19.
 /codon_start=1
 /transl_table=11
 /label=SCE29_02
 /product="putative hydrolase"
 /protein_id="CAB38877.1"
 /db_xref="GI:4490980"
 /translation="MFVLTVNGIRINYYDDAPPAGQAQNAVILLVMGSGSGRAWHLH

Query Match	1.4%	Score 52.4	DB 3	Length 26477	
Best Local Similarity	46.0%	Pred. No. 0.13			
Hatches 216	Conservative 0	Mismatches 251	Indels 3	Gaps 1	
326	cggggaagtggccagccctggccagcgcgcacccctctgtcacccggcgctgagacac	385			
Db	7425	CGGTGCCCGCCGGTGCCGGTGC	CGCGGGCGGTTCGCTCCGACCGGAGCGGCGCTGC	7484	
Qy	386	tgacacagccaccaaattctc	aactgatatgtctgcagagcaataa	tgctgcggagcaga	445
Db	7485	TGGTCAATCGAGCGCTAC	GGGAGAGTGTGGCGCGGACCGGTTCCCGGTCTACTCCGG	7544	
Qy	446	acttgaggagcagctggattgg	taccagcgctgggtgtgagcccttctgagggcgac	505	
Db	7545	TCCTCCGGGAGGACTGG	CGGGTGTGTGAGGCCGTGGAGCGCTCGGCACACTGGCGGC	7604	
Qy	506	ccagcatctcccggcgccatca	ctctcagccagcattcgtgttcgcacccggccac	565	
Db	7605	CCGCGGGCGCGCGCTGG	CGGTGTGACTACCGGGGGATGCTGAGGGGTGCACGCC	7664	
Qy	566	aggtctctctcaggccacgcg	gagagcgcgctctgtctccaaagactctctct	625	
Db	7665	AGGGGTCTTCAGCAGCGG	CACACTGGCGCGCACTTCCCGCTTCTCCAACTGTCCGGTG	7724	
Qy	626	ccgcacccccactggccaa	cgccactctggagaccgagtg---	gttccagcgctcccgcc	682
Db	7725	CGCCCGCGAGTTCGGC	ACGGCACCTGGTGTGTGGGAACTGCTCCGACGCGGGCC	7784	
Qy	683	gcaagtggagcccaacttaca	ccgcgcgcccaccatcagggcccggggacctggg	742	
Db	7785	GGCAGGTCTCTGACTGG	CGGAGACGCCCTGTACTTCCCGGACTCGACCTGGCC	7844	
Qy	743	acagctggcgccgaaggag	cggtcgtggcggggacaagaccactcaag	792	
Db	7845	AGCTCTGGATGCTCTCG	GAGAGTCTCCCGGGCGGGCGCCGCATCTGAG	7894	

RESULT	15
AK022304	
LOCUS	
DEFINITION	1669 bp mRNA
ACCESSION	Homo sapiens cDNA FLJ12422 fis, clone MAMMA1001292.
VERSION	AK022304
KEYWORDS	AK022304.1 GI:10433671
SOURCE	oligo capping; fis (full insert sequence).
	Homo sapiens Mammary gland cDNA to mRNA, clone_lib.MAMMA1
	clone:MAMMA1001292.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (sites)
AUTHORS	Isoqai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,N., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,

Search completed: November 1, 2001, 08:59:22
Job time: 8105 sec

TITLE	Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Negahari, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.
REFERENCE	NEDO human cDNA sequencing project
AUTHORS	Unpublished (2000)
TITLE	2 (bases 1 to 1669)
JOURNAL	Isogai, T. and Otsuki, T.
	Direct Submission
	Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases, Takao
	Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
	Kisarazu, Chiba 292-0812, Japan. (E-mail: genomics@hri.co.jp,
	Tel: 81-438-52-3951, Fax: 81-438-52-3952)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

```

FEATURES             Location/Qualifiers
     i. 1669
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="MAMMA1001292"
        /clone_lib="MAMMA1"
        /tissue_type="Mammary gland"
        /note="Cloning vector: pME18SFL3"
     2. 874
        /note="unnamed protein product"
        /codon_start=1
        /protein_id="BAB14007.1"
        /db_xref="GI:104333672"
        /translation="MRMEAGEAAPAGACGAGCGKGVRLNVGGTVLTTQTLCRLR
        EKPSFLRCQGEELQSDRDETGAYLIDRDPTFGPILNLRHGKLVLDKDMAEEGV
        EAEEFYNIPLIRIKDRMEKDYTVTPPKHVYRLVQCQEEETQWYVSTMSDGRWF
        EQLAVNTGSYNGSDEDAEFLCVVSKELHSTPNGLSSESRKTKSTEEQLQEQQEE
        EVEVEVEVQGVQAEADAQEGKSRPHPLRPEAEALAVRASPRPLARPOSCHPCCYKPEAP
        CEAPDILQGLGVPI"
     309 a          499 c          566 g          295 t
BASE COUNT
ORIGIN

```

Query Match	1.4%	Score	51.8	DB	89	Length	1669
Best Local Similarity	47.6%	Pred. No.	0.22				
Matches	152	Conservative	0	Mismatches	167	Indels	0
						Gaps	0
QY	165	ggactctcggtccctcgagcgcgtccaccggtggcaccatcttgccgcagaaactcgc	224				
Db	373	GGACTACACGGTCACCCAGGTCACCCCAACGATGTGTACCGGGTCTGCAGTGC	432				
QY	225	cgaggagtgccccatggaocgtggccctctacctctcacacggggactcccaaccagctgaa	284				
Db	433	GGAGGAGCTCAGCAAAATGCTCCACCATTGCTGTATGGCTGCGGCTTCGAGCAGTGGT	492				
QY	285	gcagaccactgtctccggccgctacagtgttggcggcctgcgggaaagtgcgcagccct	344				
Db	493	GAATCGGGTCTCTCTACAACTACGCAGCGAGGACCAGGCAGAGTCTCTGTGTGGT	552				
QY	345	ggcagcgcgcacccctctcttgccacggggcgctggcacactgacacagccaccaact	404				
Db	553	GTCCAAGGAGCTCCACAGCACCCCAACGGGCTGAGCTCAGTCCAGCGCGCAAAACCA	612				
QY	405	ccTcaacgtgatgtgcagagcaataagtctcggggagcagaactctgaggacgacctgaa	464				
Db	513	GAGCAGGAGGACGACTGGAGGAGCAGCAGCAGCAGGAGGAGGAGGTGGAGGAGGTGA	672				
QY	165	ttggtaccaggcgctgggtg	483				
Db	573	GGTGGAAACAGGTGCAGGTG	691				

Search completed: November 1, 2001, 08:59:22
Job time: 8105 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 06:42:12 ; Search time 159.69 Seconds
(without alignments)
14343.948 Million cell updates/sec

Title: US-09-775-181-1

Perfect score: 3648

Sequence: 1 atggagcagctgcttaacc.....gggtagttttaaaagttag 3648

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342.2	9.4	368	21	AAA42817 Human secreted exp
2	156.8	4.3	182	21	AAA41583 Human secreted exp
3	65.4	1.8	936	22	AAF58252 Oligonucleotide D1
4	65.4	1.8	936	22	AAF58254 Oligonucleotide D1
5	65.4	1.8	936	22	AAF58257 Oligonucleotide D1
6	65.4	1.8	936	22	AAF58259 Oligonucleotide D2
7	65.4	1.8	936	22	AAF58262 Oligonucleotide D2
8	65.4	1.8	936	22	AAF58255 Oligonucleotide D1
9	64	1.8	83	21	AAC08885 Human secreted pro
10	63.2	1.7	936	22	AAF58252 Oligonucleotide D1
11	63.2	1.7	936	22	AAF58254 Oligonucleotide D1

c 12	63.2	1.7	936	22	AAF58257	Oligonucleotide D1
c 13	63.2	1.7	936	22	AAF58259	Oligonucleotide D2
c 14	63.2	1.7	936	22	AAF58262	Oligonucleotide D2
c 15	63.2	1.7	938	22	AAF58255	Oligonucleotide D1
c 16	56.4	1.5	10732	21	AAA10594	Gene encoding a su
c 17	52	1.4	1035	21	AAA51616	HIV gp41 coding re
c 18	52	1.4	2466	21	AAA51614	HIV gp160 coding r
c 19	52	1.4	2547	21	AAA51615	HIV gp160 with sig
c 20	51.2	1.4	1459	21	AAC69799	Human breast tumou
c 21	51.2	1.4	3867	22	AAF28253	Human TANGO 275 op
c 22	51.2	1.4	4225	22	AAF28252	Human TANGO 275 DN
c 23	48	1.3	2571	20	AAZ08742	HIV SYNgp-160mm co
c 24	48	1.3	2571	21	AAZ93974	Human immunodefici
c 25	48	1.3	2571	22	AAZ86878	Nucleotide sequenc
c 26	46.4	1.3	2481	17	AAZ13557	Syngp160mm. Synth
c 27	46.4	1.3	2481	18	AAZ73952	HIV-1 gp160 synthe
c 28	46.4	1.3	2481	19	AAZ23291	Synthetic HIV-1 gp
c 29	45.6	1.2	2298	21	AAA51963	Modified HIV-1 Env
c 30	45.6	1.2	2298	21	AAA51964	Modified HIV-1 Env
c 31	45.6	1.2	2298	21	AAA51965	Modified HIV-1 Env
c 32	45.6	1.2	2310	21	AAA51944	Modified HIV-1 Env
c 33	45.6	1.2	2310	21	AAA51962	Modified HIV-1 Env
c 34	45.6	1.2	2316	21	AAA51945	Modified HIV-1 Env
c 35	45.6	1.2	2316	21	AAA51949	Modified HIV-1 Env
c 36	45.6	1.2	2322	21	AAA51946	Modified HIV-1 Env
c 37	45.6	1.2	2322	21	AAA51959	Modified HIV-1 Env
c 38	45.6	1.2	2322	21	AAA51960	Modified HIV-1 Env
c 39	45.6	1.2	2322	21	AAA51961	Modified HIV-1 Env
c 40	45.6	1.2	2328	21	AAA51947	Modified HIV-1 Env
c 41	45.6	1.2	2328	21	AAA51948	Modified HIV-1 Env
c 42	45.6	1.2	2352	21	AAA51967	Modified HIV-1 Env
c 43	45.6	1.2	2358	21	AAZ70447	HIV gp160.modsF162
c 44	45.6	1.2	2358	21	AAA51966	Modified HIV-1 Env
c 45	45.6	1.2	2466	21	AAZ70446	HIV gp160.modsF162

ALIGNMENTS

RESULT 1

AAA42817

ID AAA42817 standard; cDNA; 368 BP.

XX AAA42817;

AC AAA42817;

DT 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1557.

DE Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;

XX expressed sequence tag; EST; probe; chemotactic; proliferative;

KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;

KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;

KW chemilucifer; osteopathic; neuroprotective; nootropic; antipsoriatic;

KW vaccine; autoimmune disorder; anticonvulsant; antidepressant; gene therapy;

KW insulin dependent diabetes; multiple sclerosis; allergic condition;

KW lymphoid cell deficiency; asthma; myeloid cell deficiency; ulcer;

KW central nervous system disorder; osteoporosis; osteoarthritis;

KW Parkinson's disease; Huntington's disease; coagulation disorder;

KW leucopenia; thrombosis; inflammatory disorder; Crohn's disease;

KW cancer; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

```
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
XX WPI; 2000-317937/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 487; 618pp; English.
XX
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
XX Sequence 368 BP; 115 A; 78 C; 79 G; 96 T; 0 other;
XX
XX
XX Query Match 9.4%; Score 342.2; DB 21; Length 368;
XX Best Local Similarity 99.1%; Pred. No. 1e-83;
XX Matches 344; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1871 ctgctatattccatacaattagattgttcttgcctcaagacttcagctgattgattg 1930
XX || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 19 ctactatattccatacaattagattgttcttgcctcaagacttcagctgattgattg 78
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1931 tgatgctgatttgcacatactcattgactgtgacagtcaccattgggtgcttttga 1990
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 79 tgatgctgatttgcacatactcattgactgtgacagtcaccattgggtgcttttga 138
XX
XX 1991 ttccaaagtattccatacaattagattgttcttgcctcaagacttcagctgattgattg 2050
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 139 ttccaaagtattccatacaattagattgttcttgcctcaagacttcagctgattgattg 198
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2051 aggatgagtagacatgggcccagctggtatcctcactgaacagcagatcaattcagcct 2110
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 199 aggatgagtagacatgggcccagctggtatcctcactgaacagcagatcaattcagcct 258
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2111 ggagttagcacagcttggtccagagagacattcgggacgagctgaaaaactctatgcc 2170
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 259 ggagtgagcacagcttggtccagagagacattcgggacgagctgaaaaactctatgcc 318
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 2171 aactggaaatatataaagaagaagatgatcacaaaaacccccac 2217
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 319 aactggaaatatataaagaagaagatgatcacaaaaacccccac 365
XX
XX
XX RESULT 2
XX ID AAA41583
XX AAA41583 standard; cDNA; 182 BP.
```

```
XX AAA41583;
XX
XX 21-AUG-2000 (first entry)
XX
XX Human secreted expressed sequence tag SEQ ID NO:323.
XX
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX
XX Homo sapiens.
XX
XX WO2000021990-A1.
XX
XX 20-APR-2000.
XX
XX 15-OCT-1999; 99WO-US24205.
XX
XX 15-OCT-1998; 98US-0104435.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M;
XX
XX WPI; 2000-317937/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (sESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 243; 618pp; English.
XX
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
XX sequence tags (sESTs), isolated from human, mouse, xenopus and rat
XX tissue sources. The sESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
XX neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The sESTs can be used for gene
XX therapy and in vaccines. The sESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the sESTs. Proteins encoded by the sESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders
XX (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
XX disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
XX disease), tumours, bacterial, fungal or viral infections, depression and
XX psoriasis. AAA43420 to AAA43425 represent linker variants which are given
XX in the exemplification of the present invention.
XX
XX Sequence 182 BP; 46 A; 36 C; 49 G; 51 T; 0 other;
XX
```

Query Match

4.3%; Score 156.8; DB 21; Length 182;

Best Local Similarity 93.2%; Pred. No. 3.9e-33;

Matches 164; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1529 ttcttccagcaagcggtccaacgaattccatatatgactggtgacggtgtcatgagatgc 1588

Db 4 ttcggccaaagagcgctcaattgaattccatatatgactggtgacggtgtcatgagatgc 63

QY 1589 tgcagtaataactcttggttagtgtttgtttctcattggtgactggtgactctgtgtgc 1648

Db 64 tgcagtaataactcttggttagtgtttgtttctcattggtgactggtgactctgtgtgc 123

QY 1649 agaattggagaaacagatttccattatgcccaggagaaacacatccgacacctc 1704

Db 124 agaattggagaaacagatttccattatgcccaggagaaacacacccgacacacctc 179

RESULT 3

AAF58252

ID AAF58252 standard; DNA; 936 BP.

AAF58252;

24-APR-2001 (first entry)

Oligonucleotide D1835.

Electron-transfer group; ETM; mismatch; genotyping;

gene expression; ss.

Synthetic.

WO200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000WO-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

Example 6; Page 127; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 1.8%; Score 65.4; DB 22; Length 936;

Best Local Similarity 2.2%; Pred. No. 1.1e-07;

Matches 18; Conservative 436; Mismatches 357; Indels 0; Gaps 0;

QY 1411 attttgtaactttgagccaacacatttcgtgtattctcctaagatgggtctgtctc 1470

Db 1 www. 60

QY 1471 ggtttgtactgttttacggaactgtcactctcaaaacttcacagggttttgaagggtttt 1530

Db 61 www. 120

QY 1531 ctttcaagaaaggctcaacgaattccatatatgactggtgacggtgtcatgagatgctg 1590

Db 121 www. 180

QY 1591 gcagtaataactcttggttagtgtttgtttctcattggtgacttcatctgtgtgcag 1650

Db 181 www. 240

QY 1651 aatttgagaaacagatttccattatggtgcagggggaaacacccgatccatctctc 1710

Db 241 www. 300

QY 1711 aatatgtccattaccgctggactacatgacagcagttgctgaattttattctc 1770

Db 301 www. 360

QY 1771 ttgtgggtgttttctctctctatgctgacgtgcgacagtcaccatcggtatccatcgagccc 1830

Db 361 www. 420

QY 1831 cgctatatggctgtgtcagttcacaatgagctcatctctgtctatattccatacaatt 1890

Db 421 www. 480

QY 1891 agatttgtcttgcctcaagacttcagtcgtgattggtgattgtgtattttgcacat 1950

Db 481 www. 540

QY 1951 acitcatttgactgtgacagtcaccatgggtgctttgtattccaaagttttcaatca 2010

Db 541 www. 600

QY 2011 agcaataaccacgagatgattgtctacagaagcatatgaggtgagctagacatgggc 2070

Db 601 www. 660

QY 2071 cgatctggtacctacctgaacagcagtcataatcagcctggagtgagcacagcttgat 2130

Db 661 www. 720

QY 2131 ccagaggacattcgggacgagctgaaaaaactctatgcccaactggaaatatataaaga 2190

Db 721 www. 780

QY 2191 aagaagatgatcacaaacaccccccacctcc 2221

Db 781 www. 811

RESULT 4

AAF58254

ID AAF58254 standard; DNA; 936 BP.

XX AAF58254;

XX AAF58254;

DT 24-APR-2001 (first entry)

XX Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;

XX gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.


```
QY 1591 gcagtaatactctggtagtgatttgggttctctcattgagctgacctcattctgtgtgcccag 1650
Db 181 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 240
QY 1651 aatttgagaaacagatttcaacttattggccagggggaaacacatccgatccatctcttc 1710
Db 241 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 300
QY 1711 aatgtgctcattgacccgctgggactacatgacagcagctgtgtgattttttctctc 1770
Db 301 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 360
QY 1771 ttgtgggggtgttattctctgctatgctgagtcgagacagtcacatccatccatgagccc 1830
Db 361 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 420
QY 1831 cgctatatggctgtgagttcacaatgagctcattctctctctctctctctctctctcacaatt 1890
Db 421 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 480
QY 1891 agatttcttctgctcaagactcagctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1950
Db 481 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 540
QY 1951 actcatttgactgtgacagtcacacattgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2010
Db 541 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 600
QY 2011 agcaataaccacagagatgattgtctacagaagcagctatgagtgatgagtgatgagtgagc 2070
Db 601 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 660
QY 2071 cgatctggatctacctgaacagcagctatcaattcagcctgagtgagcagcagctgtgtgt 2130
Db 661 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 720
QY 2131 ccagaggacattcgggagcagctgaaacacacacacacacacacacacacacacacacacacac 2190
Db 721 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 780
QY 2191 aagaagatgatcacaaacacacacacacacacacacacacacacacacacacacacacacac 2250
Db 781 www.ggggggggggggggggggggggggggggggggggggggggggggggggggggggggggg 840
```

RESULT 6

AAF58259

AAF58259 standard; DNA; 936 Bp.

AAF58259;

24-APR-2001 (first entry)

Oligonucleotide D2004.

Electron-transfer group; ETM; mismatch; genotyping;
gene expression; ss.

Synthetic.

W0200107665-A2.

01-FEB-2001.

26-JUL-2000; 2000W0-US20476.

26-JUL-1999; 99US-0145695.

17-MAR-2000; 2000US-0190259.

(CLIN-) CLINICAL MICRO SENSORS INC.

Umek RM;

XX

DR

XX WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface

PS Example 6; Page 128; 159pp; English.

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.

XX Sequence 936 Bp; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 1.8%; Score 65.4; DB 22; Length 936;

Best Local Similarity 2.2%; Pred. No. 1.le-07;

Matches 18; Conservative 436; Mismatches 357; Indels 0; Gaps 0;

QY 1411 atttgtactttgagccaagcacatttgcgtgtattctcctaagatgggctgctctc 1470

Db 1 www.ggg 60

QY 1471 ggtttgtctactgtttacggaactgtcactctcaaaccttcacaggttttgaagtggttt 1530

Db 61 www.ggg 120

QY 1531 ctftcagaaacggtcacaagattccatatatgactggcgagcggtcatgagatgctg 1590

Db 121 www.ggg 180

QY 1591 gcagtaatactctgt 1650

Db 181 www.ggg 240

QY 1651 aatttgagaaacagatttcaacttattggccagggggaaacacatccgatccatctcttc 1710

Db 241 www.ggg 300

QY 1711 aatgtgctcattgacccgctgggactacatgacagcagctgtgtgtgtgtgtgtgtgtgtgt 1770

Db 301 www.ggg 360

QY 1771 ttgtgggggtgttattctctgctatgctgagtcgagacagtcacatccatccatgagccc 1830

Db 361 www.ggg 420

QY 1831 cgctatatggctgtgagttcacaatgagctcattctctctctctctctctctctctcacaatt 1890

Db 421 www.ggg 480

QY 1891 agatttcttctgctcaagactcagctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1950

Db 481 www.ggg 540

QY 1951 actcatttgactgtgacagtcacacattgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2010

Db 541 www.ggg 600

QY 2011 agcaataaccacagagatgattgtctacagaagcagctatgagtgatgagtgatgagtgagc 2070

Db 601 www.ggg 660

QY 2071 cgatctggatctacctgaacagcagctatcaattcagctgagtgagtgagtgagtgagtgag 2130

Db 661 www.ggg 720

QY 2131 ccagaggacattcgggagcagctgaaacacacacacacacacacacacacacacacacacacac 2190

Db 721 www.ggg 780

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 07:46:12 ; Search time 66.1 Seconds
(without alignments)
10447.912 Million cell updates/sec

Title: US-09-775-181-1
Perfect score: 3648
Sequence: 1 atggagcagtgcttacc.....gggtagttttaaaagttag 3648
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	62	1.7	7218	1	US-08-232-463-14
2	48.4	1.3	7218	1	US-08-232-463-14
3	46.4	1.3	2481	1	US-08-324-243-35
4	46.4	1.3	2481	1	US-08-532-390-35
5	46.4	1.3	2481	3	US-08-717-294-35
6	46.4	1.3	2481	5	PCT-US95-11511-35
7	44.8	1.2	2502	1	US-08-073-384C-7
8	44.8	1.2	2502	1	US-08-254-359A-7
9	44.8	1.2	2502	1	US-08-483-043-7
10	44.8	1.2	2502	1	US-08-481-238-7
11	44.8	1.2	2502	2	US-08-471-066B-7
12	44.8	1.2	2502	2	US-08-484-956-7
13	44.8	1.2	2502	2	US-08-757-653-7
14	44.8	1.2	2502	2	US-08-599-491-7
15	44.8	1.2	2502	2	US-08-756-386-7
16	44.8	1.2	2502	2	US-08-823-516-7
17	44.8	1.2	2502	3	US-08-682-853A-7
18	44.8	1.2	2502	3	US-08-759-038-7
19	44.8	1.2	2502	3	US-08-758-314-7
20	44.8	1.2	2634	3	US-08-911-853-30
21	44.8	1.2	2634	3	US-08-479-409-30
22	44.8	1.2	17612	3	US-08-911-853-29
23	44.8	1.2	17612	4	US-09-479-409-29
24	42.8	1.2	771	2	US-08-972-008-3
c 25	42.8	1.2	2525	2	US-08-972-008-1
c 26	42	1.2	44377	2	US-08-804-227C-7
27	42	1.2	44377	2	US-08-804-198-1

29	41.8	1.1	2496	1	US-08-073-384C-2	Sequence 2, Appli
29	41.8	1.1	2496	1	US-08-254-359A-2	Sequence 2, Appli
30	41.8	1.1	2496	1	US-08-483-043-2	Sequence 2, Appli
31	41.8	1.1	2496	1	US-08-481-238-2	Sequence 2, Appli
32	41.8	1.1	2496	2	US-08-471-066B-2	Sequence 2, Appli
33	41.8	1.1	2496	2	US-08-484-956-2	Sequence 2, Appli
34	41.8	1.1	2496	2	US-08-757-653-2	Sequence 2, Appli
35	41.8	1.1	2496	2	US-08-599-491-2	Sequence 2, Appli
36	41.8	1.1	2496	2	US-08-756-386-2	Sequence 2, Appli
37	41.8	1.1	2496	2	US-08-823-516-2	Sequence 2, Appli
38	41.8	1.1	2496	3	US-08-682-853A-2	Sequence 2, Appli
39	41.8	1.1	2496	3	US-08-759-038-2	Sequence 2, Appli
40	41.8	1.1	2496	3	US-08-758-314-2	Sequence 2, Appli
41	40.6	1.1	4776	2	US-08-458-568A-11	Sequence 11, Appl
c 42	40.4	1.1	12001	1	US-08-458-568A-11	Sequence 11, Appl
43	40.4	1.1	1524	2	US-08-716-942-24	Sequence 24, Appl
c 44	40.4	1.1	1771	2	US-08-533-669A-7	Sequence 7, Appli
c 45	40.4	1.1	1771	2	US-08-511-872-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-Fls
; US-08-232-463-14

CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 04-JUN-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-073-384C-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;

Best Local Similarity 46.8%; Pred. No. 0.038;

Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

Qy 413 tgatgtcagagcaataagtcgaggagcagaacttcgaggagcacttgattggtacc 472
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTTGGCCACCTCTCCGGGGACG 1879
Qy 473 aggcgtgtgtgagccttctgagggcgagcccgagcactctccggcgccacacct 532
Db 1880 AGACCTGATCCGGGTCTTCCAGGAGGGGAGGACATCCACACCCAGCCAGCTGGA 1939
Qy 533 tcagcaccgattcgtctccgacccggccacaggttcttccacagccacgcgag 592
Db 1940 TGTTCCGGCGTCCCGCGAGCCGTGACCCCTGATGCGCGCGCGCCCAAGACCATCA 1999
Qy 593 agagccgacatctgctccagacccgtctctctccgacccacacacacacacacac 652
Db 2000 ACTTCGGGGTC---CTCTACGGCATGTCCGCCACCCACCCCTCTCCAGGAGCTTCCCAATCC 2056
Qy 653 tggagaccgagtggttccacggcctccgagcagtgaggccacacacacacacacac 712
Db 2057 CCTACGAGGAGCGGTGGCTTCTTCCAGGCGCTACTTCCAGAGCTTCCCAAGGTGCGG 2116
Qy 713 gcccaatcagggggccgggcttcctccgacagctggcgccgcaaggggctcgcg 772
Db 2117 CCTGATTGAGAGACCCCTGGAGGAGGCGAGGGCGGGGTACGTGGAGACCCCTCTTCG 2176
Qy 773 gggacaagaccactt 788
Db 2177 GCCCGCGCGCTACGT 2192

RESULT 8

US-08-254-359A-7

Sequence 7, Application US/08254359A

Patent No. 5614402

GENERAL INFORMATION:

APPLICANT: DAHLBERG, JAMES E.

APPLICANT: LYANICHEV, VICTOR I.

APPLICANT: BROW, MARY ANN D.

TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
TITLE OF INVENTION: DNA POLYMERASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,359A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-254-359A-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;

Best Local Similarity 46.8%; Pred. No. 0.038;

Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

Qy 413 tgatgtcagagcaataagtcgaggagcagaacttcgaggagcacttgattggtacc 472
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTTGGCCACCTCTCCGGGGACG 1879
Qy 473 aggcgtgtgtgagccttctgagggcgagcccgagcactctccggcgccacacac 532
Db 1880 AGACCTGATCCGGGTCTTCCAGGAGGGGAGGACATCCACACCCAGCCAGCTGGA 1939
Qy 533 tcagcaccgattcgtctccgacccggccacaggttcttccacagccacgcgag 592
Db 1940 TGTTCCGGCGTCCCGCGAGCCGTGACCCCTGATGCGCGCGCGCCCAAGACCATCA 1999
Qy 593 agagccgacatctgctccagacccgtctctccgacccacacacacacacacacac 652
Db 2000 ACTTCGGGGTC---CTCTACGGCATGTCCGCCACCCACCCCTCTCCAGGAGCTTCCCAATCC 2056
Qy 653 tggagaccgagtggttccacggcctccgagcagtgaggccacacacacacacacac 712
Db 2057 CCTACGAGGAGCGGTGGCTTCTTCCAGGCGCTACTTCCAGAGCTTCCCAAGGTGCGG 2116
Qy 713 gcccaatcagggggccgggcttcctccgacagctggcgccgcaaggggctcgcg 772
Db 2117 CCTGATTGAGAGACCCCTGGAGGAGGCGAGGGCGGGGTACGTGGAGACCCCTCTTCG 2176
Qy 773 gggacaagaccactt 788
Db 2177 GCCCGCGCGCTACGT 2192

RESULT 9
US-08-483-043-7
; Sequence 7, Application US/08483043
; Patent No. 5691142
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,043
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,384
; FILING DATE: 04-JUN-1993
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cartoli, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-043-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;
Qy 413 tgatgtcgagagcaataagtcgcgggagcagaacttcgaggacacacctggaattggtacc 472
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGTCCGGGTCTCTGGCCACCTCTCTCCGGGACG 1879
Qy 473 aggcgctggttgagccttctgagggcgagccagcatctcccgggcgccatcacct 532
Db 1880 AGAACCCTGATCCGGGTCTTCCAGGAGGGAGGAGCATCCACACAGACCCGCGCTCGA 1939
Qy 533 tcagcaccgattcgctgctccgcaccgcgccacaggttctctccagggcgccgaggg 592
Db 1940 TGTTCGGCGTCCCCCGGAGCGCTGGACCCCTGATGCGCGCGGCGCCAGACCATCA 1999
Qy 593 agagccgcatctgctccaagacctgtctctctccgacccacacacacacacacacac 652
Db 2000 ACTTCGGGGTC---CTCTACGCGATGTCCGCCACCGCTCTCCAGGAGCTTGCCATCC 2056
Qy 653 tggagaccgagtggttccacggcctccggcgcaagtggagggcccaacttacacggcg 712
Db 2057 CCTACGAGGAGCGGTGGCCCTTCATTGAGCGCTACTTCCAGAGCTTCCCCAAGGTGCGG 2116

Qy 713 gcccaatcagggcccccggggccctgggcccacagctggcgcgcaaggacgggctcgcg 772
Db 2117 CTTGATTCAGAGACCCCTGGAGAGGGGAGGAGCGGGGGTACGTGGAGACCCCTCTTCG 2176
Qy 773 gggacaagagccactt 788
Db 2177 GCCGCGGGCTACGT 2192

RESULT 10
US-08-481-238-7
; Sequence 7, Application US/08481238
; Patent No. 5795763
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,238
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-481-238-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.038;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;
Qy 413 tgatgtcgagagcaataagtcgcgggagcagaacttcgaggacacacctggaattggtacc 472
Db 1820 TGGTGGCCCTGGACTATAGCCAGATAGAGTCCGGGTCTCTGGCCACCTCTCTCCGGGACG 1879
Qy 473 aggcgctggttgagccttctgagggcgagccagcatctcccgggcgccatcacct 532
Db 1880 AGAACCCTGATCCGGGTCTTCCAGGAGGGAGGAGCATCCACACAGACCCGCGCTCGA 1939
Qy 533 tcagcaccgattcgctgctccgcaccgcgccacaggttctctccagggcgccgaggg 592
Db 1940 TGTTCGGCGTCCCCCGGAGCGCTGGACCCCTGATGCGCGCGGCGCCAGACCATCA 1999
Qy 593 agagccgcatctgctccaagacctgtctctctccgacccacacacacacacacacac 652
Db 2000 ACTTCGGGGTC---CTCTACGCGATGTCCGCCACCGCTCTCCAGGAGCTTGCCATCC 2056

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2001, 06:45:12 ; Search time 1878.46 Seconds
(without alignments)
18357.596 Million cell updates/sec

Title: US-09-775-181-1
Perfect score: 3648
Sequence: 1 atgggagccatggcttacc.....gggatagttttaagtgtag 3648

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est33.*
26: gb_est34.*
27: gb_est35.*
28: gb_est36.*
29: gb_est37.*
30: gb_est38.*
31: gb_est39.*
32: gb_est40.*
33: em_estba.*
34: em_estfun.*
35: em_esthum1.*
36: em_esthum2.*
37: em_esthum3.*
38: em_esthum4.*
39: em_esthum5.*
40: em_esthum6.*
41: em_esthum7.*
42: em_esthum8.*
43: em_esthum9.*
44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estomi.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est41.*
111: gb_est42.*
112: gb_est43.*
113: gb_est44.*
114: gb_est45.*
115: gb_est46.*
116: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	380.8	10.4	508	AQ764784	HS_3118_A
2	372.4	10.2	534	AQ837108	HS_5464_A
3	336.2	9.2	543	AQ728799	HS_5464_B
4	222.6	6.1	681	AU171941	AU171941
5	219.2	6.0	280	BB354926	BB354926
6	194	5.3	288	BB429778	BB429778
7	172.6	4.7	895	CNS032UA	AL225307 Tetraodon
8	166.8	4.6	632	CNS02NOT	AL205670 Tetraodon
9	165	4.5	961	CNS022LV	AL178348 Tetraodon
10	108.2	3.0	592	AQ881884	HS_5304_A
11	101.2	2.8	877	CNS03H19	AL243702 Tetraodon
12	95.4	2.6	329	A2628814	LM0481D17
13	93.2	2.6	1030	CNS03EM8	AL240569 Tetraodon
14	87.2	2.4	988	CNS04BT9	AL283590 Tetraodon
15	86	2.4	633	CNS04FKJ	AL288460 Tetraodon
16	86	2.4	921	CNS03PB5	AL254426 Tetraodon
17	82	2.2	888	CNS03YL7	AL266452 Tetraodon
18	80.8	2.2	807	CNS0317B	AL223184 Tetraodon
19	80.6	2.2	957	CNS020UA	AL207163 Tetraodon
20	80.4	2.2	822	CNS02FTA	AL195463 Tetraodon
21	78.2	2.1	899	CNS032UB	AL225308 Tetraodon
22	78	2.1	483	A2903235	RPCT-24-1
23	77	2.1	1096	CNS05CMB	AL331292 Tetraodon
24	75.6	2.1	379	CNS03EM7	AL240568 Tetraodon
25	73.2	2.0	1012	CNS03Y7G	AL265957 Tetraodon
26	72.8	2.0	446	A2870959	AL2870959 2M0183F08
27	63.6	1.7	817	CNS05ATS	AL328969 Tetraodon
28	63.4	1.7	925	CNS0091P	AL053013 Drosophil
29	61.8	1.7	591	AQ764779	HS_3118_A
30	61.6	1.7	1043	CNS020JP	AL175678 Tetraodon
31	61	1.7	902	CNS006QP	AL065804 Drosophil
32	60.4	1.7	923	CNS0091P	AL053013 Drosophil
33	58.4	1.6	518	A2823042	2M0096N15
34	58.2	1.6	1101	CNS012S8	AL101954 Drosophil
35	57.4	1.6	839	CNS004NB	AL054280 Drosophil
36	56	1.5	932	CNS0072Q	AL066742 Drosophil
37	55	1.5	844	CNS0052P	AL056652 Drosophil
38	55	1.5	972	CNS01XPU	AL172011 Tetraodon
39	54.4	1.5	1101	CNS00397	AL063912 Drosophil
40	53	1.5	1100	CNS016KD	AL106855 Drosophil
41	51.6	1.4	1101	CNS017SY	AL108460 Drosophil
42	51.2	1.4	608	BF688560	602184874
43	51.2	1.4	918	AQ895329	AQ895329
44	51	1.4	839	CNS004NB	AL054280 Drosophil
45	50.6	1.4	817	BE558893	HV_CEB002

ALIGNMENTS

RESULT 1	AQ764784	508 bp	DNA	GSS	28-JUL-1999
LOCUS	HS_3118_AL_C10_MR CIT	Approved Human Genomic Sperm Library D Homo			
DEFINITION	sapiens genomic clone Plate-3118 Col-19 Row=E, DNA sequence.				
ACCESSION	AQ764784				
VERSION	AQ764784.1	GI:5642900			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 508) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and				

JOURNAL
MEDLINE
COMMENT

scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3118 row: E column: 19
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 508.

FEATURES

source

Location/Qualifiers
1..508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3118 Col=19 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 166 a 129 c 109 g 98 t 6 others
ORIGIN

Query Match

10.4%; Score 380.8; DB 233; Length 508;
Best Local Similarity 96.4%; Pred. No. 9.5e-95;
Matches 407; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 2143 cgggacgagctgaaacactctatgcccaactggaatatataaaagaagaagatgc 2202
|||||
DB 88 CAGGACGAGCTGAAAAAATCTATGCCCACTGGAAATATATAAAGAGAGATGATC 147
QY 2203 acaaaacacccccacctccagaaaaagcggctgcgaagaaggcctcagtgcttaac 2262
|||||
DB 148 ACAAAACAACCCCACTCCAGAAAAGCGGTGCTCGAAGAAGAGCGCTTANGTCGTCCATC 207
QY 2263 atagacgacattacggagatcccgagacacagtcagcggcagtgcttaagagagacaag 2322
|||||
DB 208 ATGAGACGATTCGAGAGATCCCGAGACAGTACGCCGCGCAGTCTCTAAAGAGGAGCAAG 267
QY 2323 gagggcgccgacctggcagagcgaaggcactgccctcaccaggaagacccccagag 2382
|||||
DB 268 GAGGCGCGGACCATGGCAGACGCCAAGGCACTGCCCTCATCAGAGAACTCCAGAG 327
QY 2383 tcttcagggaacacaggggaatccaaaggagagaccctgaaaaaccagtgcttctcactc 2442
|||||
DB 323 TCTTCATGGAACACATGGAATCCAAAGGAGAGACCCCTG-ATAACCGAGTCTTCTCACTC 386
QY 2443 aagaatcccaacacacttatgaccagtgagagacaaacgaagagtcagtagccta 2502
|||||
DB 387 AAGAAATCCACACACATTTATGACACGTGAGAGACCANAGAGAGTCCAGTAGCCTA 446
QY 2503 cccacagaaagcagagagagagacacagaaatccacactggaatcccttcgggt 2562
|||||
DB 447 CCCACAGAAAGCCATGAGGAGAGACACACAG-AAATTNCACACTGGAAATCCNTGCGGT 505
QY 2563 aa 2564
||
DB 506 AA 507

RESULT 2

AQ837108

LOCUS

DEFINITION

ACCESSION

VERSION

AQ837108 554 bp DNA GSS
HS_5464_AL_H02_T7A RPCT-11 Human Male BAC Library Homo sapiens
genomic clone Plate-1040 Col-3 Row=O, DNA sequence.
AQ837108
AQ837108.1 GI:5806982

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 554)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1040 row: O column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 554.
Location/Qualifiers
1..554
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1040 Col-3 Row-O"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 182 a 147 c 119 g 102 t 4 others
ORIGIN

Query Match 10.2%; Score 372.4; DB 234; Length 554;
Best Local Similarity 92.0%; Pred. No. 2.2e-92;
Matches 414; Conservative 0; Mismatches 33; Indels 3; Gaps 2;
QY 2157 aaactctatgcccaactggaataataaaagaagaagatgatcacaaaccccca 2216
DB 93 ARAACTCTATGCCCACTGGAATATATAAAGAAGAGATGATCACAAACCCCA 152
QY 2217 cctcagaaaaagcgtctcgaagaaggcctaggtcttcctcatcatgagacgattac 2276
DB 153 CCTCAGAAAAAGCGTGCTCGAAGAAAGCGCTANGTCGTCATCATGAGACGATTAC 212
QY 2277 ggagatccagagacagtcagcggcagctgtctctaaagagacaaggagggcgccagca 2336
DB 213 GGAGATCCAGAGACGTCAGCGCGCAGTGTCTCTAAAGAACAAGAGGGCGCGACCA 272
QY 2337 ttgcacagccaaagcactgcctcatcatggaagaacccccagagcttccagggaacac 2396
DB 273 TGGCAGACCCAAAGCACTGCCTCATATGAGAAACCCCGAGAGTCTTCAGGAACAC 332
QY 2397 agggaaatccaaaggaggagaccctgaaacacgagctcttctcactcaagaatccacag 2456
DB 333 AGGGAATTCAGAGAGAGACCTGAAAAACCGAGTCTTCTCACTCAAGAAATCCACAG 392
QY 2457 cacttataccacgtgagagaccacaaacggaaggtccagtagctaccacagaaagcca 2516
DB 393 CACTTATGACCACGTGAGAGACCAACCCAGAGTTTACCTTACCACCAAGATGTC- 451

QY 2517 agagagaggagacacacagaaattccacactggaatccctgctgggtataaaataacaca 2576
DB 452 AGAAGAGGAGACTACAGAAATCCACACTGGAATCTCTGTCNGGTAAAACTTA--CCC 509
QY 2577 aaaaataaagaagacagcagcgagctgagtc 2606
DB 510 GCTATCTATAGAAGACAGCGAGGCTTATTC 539

RESULT 3
QY 28799 543 bp DNA GSS 15-JUL-1999
LOCUS HS_5464_B2_A11_77A_RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate-1040 Col-22 Row-B, DNA sequence.
ACCESSION AQ728799
VERSION AQ728799.1 GI:5500351
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 543)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1040 row: B column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 543.
Location/Qualifiers
1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1040 Col-22 Row-B"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 168 a 139 c 124 g 107 t 5 others
ORIGIN

Query Match 9.2%; Score 336.2; DB 232; Length 543;
Best Local Similarity 87.4%; Pred. No. 2.7e-82;
Matches 368; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 2146 gacgagctgaaaaactctatgcccaactggaatataataaaagaagaagatgatcaca 2205
DB 110 GACGAGCTGAAAACACTCTATGCCCACTGGATGTGTTTCAAGAAAGAGATGATCTCA 169
QY 2206 aaaaacccccacctccagaaaaagcgtctcgaagaaggccttaggtcgttccatcatg 2265
|||||

	79	a	65	c	62	g	74	t	
Query Match			6.0%		Score	219.2;	DB	132;	Length 280;
Best Local Similarity			86.4%		Pred. No.	8.9e-50;			
Matches	242;	Conservative	0;	Mismatches	38;	Indels	0;	Gaps	0;
QY	1897	gtcttgccctcaagacttcagtcctgattggtgattgattgctgtattttgcacatactcat	1956						
Db	1	GTCTAGCTTTCGAGATTTCAGSCCCGATTGGATGCTGATGCTGACTTTGCACATTCGTTAC	60						
QY	1957	ttgacttgacagtcacacattgggttgccttttgcattccaaagttttcacattcaagaacat	2016						
Db	61	TTGACTGTGACAGTTACCACTGGCGTCTCTAATTCCAAAGTTTTCACATTCACGCAAT	120						
QY	2017	aaccacagagatgatattgctacagaagcatatgagatgagctagacatgggcccgatct	2076						
Db	121	AATCCCGCAGAAACACATTGTCAGGAAGCATATGAAGATGAGCTGGACATGGTTCGCTCT	180						
QY	2077	ggatcctaccctgaacagcagatcaattcaacctggagtgagcacagcttgggtccagag	2136						
Db	181	GGATCCTACCTGAAACAGCAGTATCAATTCAAGCTGGAGTGAACACAGCCCTAGATCCAGAA	240						
QY	2137	gacattcgggacgagctgaaaaaactctatgcccaactcg	2176						

FEATURES	
SOURCE	

Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 632)
Roest-Croillius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 632)
Genoscope.
Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.

```

      TURES source
            1..632
              /organism="Tetraodon nigroviridis"
              /db_xref="taxon:99883"
              /clone_1="151N10"
              /clone_lib="C"
              /notes="Genoscope sequence ID : COAGI51DG05LP1-end : T7"
BASE COUNT   99 a    206 c    219 g    99 t     9 others
ORIGIN
Query Match          4.6%; Score 166.8; DB 220; Length 632;
Best Local Similarity 59.2%; Pred. No. 5e-35;
Matches 334; Conservative 3; Mismatches 221; Indels 6; Gaps 3;

QY 341 ccctggcagcgccaccctcttgcacggg-gctggacacctgacacgcgccac 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CACCGCGCGCCGCAACCGCTCGGTTCAGCGCGSGTGCGACGGCTGCACGCCGCC 546

QY 400 aaacttcctaactgatgtcagagaataaagtctcgaggagagaaacttgacggac 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 AACPTCTCATCGGTATCTTCGAAGCCAACAGGTTCGGAGGAGAGA--CTGCGCGCCAA 488

QY 460 ctggattgttaccaaggcgctggtgtggagctctctgagggcgagcccagcatctccgg 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 ATGGAGTGTATCCACGCTCTGTTCSGAGGATCCTTGSAGGGGGATTCACAGATCCACGG 428

QY 520 gcggccaactccttagcacgcagtctgcgcacgggcccccacaggtcttcctccag 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 GCTGTGTCTACTTGGGGCCGATGC---GTCTCGTGGGGRACCCCGGGTGCTGCTCCAG 371

QY 580 gccacgcgagagagagcgcactctctccagacctgtctctctccgccccaccctg 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GCACACAGGCGCGCCACAGCAGATGCTGCTCAGAGACTGTTCACAGCTGGCGCGCGCTC 311

QY 640 gccaacgceactctgagacccgagtggttcacgcgctccgcgcgaagtggagggccccac 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 MACAACCGCACCGCCACACGGGCTGTGTACCACGACGCCAAGAGAGGCCGAGAGAGCCC 251

QY 700 ttacacgcgcgcgcgcacaatcagggcccggggctggcgcacagctggcgcgcaag 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 RGCTTCGGGAAGAGGTTCCTGACCCAGAACCCAGCCCTAGATTCTGTGGCAGGGGGAG 191

QY 760 gacgggctcggcggggacaaaagaccacttaagtgtctctccgecttatctgagtgcgag 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 CCAGAGTGTGTCCCGGACAGRGCGCACGTGAGGTGGTTCGCGCGCCCTACCTGGAGTTGGC 131

QY 820 aecgggagttacaagcccggtggtggttaactcttctctccatctacgggttcgag 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 GAGGGCGTCHGCTCCACAGCTGGCTCTCTGACCTTATCRGCTGCCCTCTACGGCCTGAAG 71

QY 880 cctaaactggtctcccggaattcagg 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


	Query Match	2.48;	Score 86;	DB 221;	Length 633;
	Best Local Similarity	73.38;	Pred. No. 1.7e-12;		
	Matches 110;	Conservative	0;	Mismatches 40;	Indels 0;
				Gaps	0;
Qy	1997	agtttcacattcaagcaataaccacagagatgattgtcacagagcatataggagt	2056		
Db	343	AGTTCTCTTTGCGCGGACCCCATGTGGAGACGACATGGCCCTCGAGGCCCTACGAGACG	402		
Qy	2057	agctagacatgggcccgatctggatcctcaacctgaacagcagtcataatcagccctggagtg	2116		
Db	403	AGCTGGACATGGGCGCGCTCCGGGTCTACCTCAACACGACGATCACGTCGCGCTGGAGCG	462		
Qy	2117	agcacagctctggattccagagagacattcggg	2146		
Db	463	AGCACGCTTGGATCCCGGAACACATTCGGG	492		

Search completed: November 1, 2001, 09:12:23
Job time: 8831 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 09:12:27 ; Search time 45.13 Seconds
(without alignments)
1632.133 Million cell updates/sec

Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGNMAYPLLLCLLLAQLGLG.....LSANKIAGPRKEEINDSPKV 1215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*

1:	/SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223.5	3.5	1199	13	AA25080
2	220	3.4	1219	18	AAW25763
3	212	3.3	1197	20	AAV28562
4	210	3.3	1171	20	AAV28563
5	198	3.1	1180	16	AAW64253
6	198	3.1	1397	20	AAV49134
7	197.5	3.1	1212	20	AAW28564
8	193.5	3.0	1212	16	AAW64254
9	193.5	3.0	1394	20	AAV49129
10	193	3.0	1194	14	AAW2199
11	190	3.0	1056	18	AAW25765

12	186	2.9	1056	18	AAW25764	Amino acid sequenc
13	183.5	2.9	1276	20	AAV49127	phCAR/hmglur2*Gq15
14	180	2.8	906	13	AAW25081	GLU-G-R subtype 1b
15	179.5	2.8	3266	21	AAW42491	Human ORFX ORF2255
16	178	2.8	571	19	AAW55064	Streptococcus pneu
17	177	2.8	906	16	AAW55064	Human mglur1b, Ho
18	176	2.8	917	20	AAW49126	Chimeric hCAR/hmgl
19	176	2.8	1035	20	AAV49128	Chimeric hCAR/hmgl
20	172.5	2.7	982	19	AAW71632	Human myosin I-cha
21	172.5	2.7	1079	19	AAW54847	Rat kidney calciu
22	172.5	2.7	1079	19	AAW38275	Rat kidney cell ca
23	172.5	2.7	1079	20	AAV41781	Rat parathyroid ca
24	172.5	2.7	1079	20	AAW94928	Rat kidney extrac
25	172.5	2.7	1079	20	AAW89566	Rat parathyroid ca
26	172.5	2.7	1079	21	AAV51828	Rat calcium recept
27	172.5	2.7	1418	20	AAV49131	mglur8/Car*Galphag
28	170	2.7	877	16	AAW64255	Human mglur5c, Ho
29	169	2.6	554	19	AAV20763	Human neurofilamen
30	169	2.6	974	21	AAV45000	Human calcium sens
31	168.5	2.6	1257	22	AAW66926	RBPL protein, Hom
32	168	2.6	1863	21	AAV77818	BRCA1 protein sequ
33	167.5	2.6	682	21	AAG46506	Arabidopsis thalia
34	167.5	2.6	724	21	AAG46505	Arabidopsis thalia
35	167.5	2.6	778	21	AAG46504	Arabidopsis thalia
36	167	2.6	1863	18	AAW23286	Human breast and o
37	166	2.6	1114	21	AAW08517	Amino acid sequenc
38	166	2.6	1596	18	AAW31347	Rat tumour suppres
39	164	2.6	1965	19	AAW69165	Streptococcus pneu
40	163.5	2.6	480	21	AAV82317	Human protein tran
41	163.5	2.6	877	16	AAW82657	Human mglur3, Hom
42	163.5	2.6	901	21	AAV45001	Human calcium sens
43	163	2.6	718	21	AAW42569	Arabidopsis thalia
44	163	2.6	738	21	AAW42568	Arabidopsis thalia
45	163	2.6	763	21	AAW42567	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAW25080

ID AAR25080 standard; Protein; 1199 AA.

XX AAR25080;

XX 04-JAN-1993 (first entry)

XX GLU-G-R subtype 1a.

XX Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;

KW neurotransmitter; glutamate; neuronal excitation; clone 45-A.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 98

FT Modified-site /note= "N-glycosylation"

FT Modified-site 223

FT Modified-site /note= "N-glycosylation"

FT Modified-site 397

FT Modified-site /note= "N-glycosylation"

FT Modified-site 515

FT Modified-site /note= "N-glycosylation"

FT Domain 594..614

FT Domain /label= Transmembrane_domain

FT Domain 630..650

FT Domain /label= Transmembrane_domain

FT Domain 657..677

FT Domain /label= Transmembrane_domain

FT Domain 707..727

FT Domain /label= Transmembrane_domain

FT Modified-site 747

FT Modified-site /note= "N-glycosylation"

Best Local Similarity 18.5%; Pred. No. 1.1e-06;		Matches 177; Conservative 126; Mismatches 402; Indels 252; Gaps 35;	
QY	219	ETWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPPYLECE	273
Db	332	dvkwdyylklrpetnlrnp-----wfgewqhrfqrlegfaqenskyntk-----cn	381
QY	274	NG-----SYKPGWLW-TLSSAIYGL---QPNLVPEFRGV-----MKVDINLQ	311
Db	382	ssltlrthvqskmgfvinaiysmayghnmqmslcpgyagldcmkpidgrkllidslm	441
QY	312	KVDIDQCSDDGWFSGTHKCHLNNSCEMPTKGLGVLGAYECICKAGFYHPGVLVPV-----	366
Db	442	ktnftvgsgdmllfdengdspgyrlylmnfkeng---kdyfdyinvgswdngelkmddev	498
QY	367	-----NNFRRRGPDOHI-SGSTKDVSEE---AYVCLPCREG-----CPFC-----A	403
Db	499	wskknllirsvcepecekgqkvrkgevcswctctpckeneyvfydeytcckacqlgsupt	558
QY	404	DD-SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVYHFRKAKSIRASGLILLE	456
Db	559	ddltgdcdlipqylrvgdpepiaavvfacglilatitfvtfviifdyrdtpvkvssrelcy	618
QY	457	TILFGSELLLYFPVILYFEPSTFRFCILLRWARLLGFATVGTWTLKLRVLKV-----	509
Db	619	illagiclgylctfcilakpkqicylqrligispamsysalvtktnriarilagskkk	678
QY	510	-----FLSRTAQRIPYMTGGVRMRLAVILLVWPFILIGWTFSSVCNLEKQISL----	558
Db	679	icckkprfmsacq-----lviafili-----ciglgiilvafime	714
QY	559	---IGOGKTSDDLIFNMCLDRWDYMTAVA-EPLFLLMGVLYCYAVRTVPSAFHEPRYMA	614
Db	715	ppdimbdpsirevyltntnlgvvtplgyngllilscftvafktrnvpantneakyla	774
QY	615	VAVHNELIISAHTTRFVLASRLQSDWMLMXYFAUTHTVTVITIGLLIPK-----FSHS	670
Db	775	fmyttciilwafviyfyf-----gsnykiltmcfavsatsvalgcmfvpkyviiakp	828
QY	671	SNPRDIIATE-AYEDELMDGRSGYL--NSSINSAMSEHSLDPDIRDELKKL---YAQ	724
Db	829	ernvrsaftstvmnhvgdgksssaarssslvnlwkrsgsgctirykdrriagkhs	888
QY	725	LEIYKRKKMITNPHLQKRCSSKGL-----GRSIMRR-----I	758
Db	889	iecfpkpgmgnggratmssngksvtwaqnekstrgqhlwqrlsvhinkkenpqtavi	948
QY	759	TETPETVSRQCKEDKEGADHTAGK-----TALIRKNPPSSSGWTGKSKEETLKN	809
Db	949	kpfpkate---arggagaggggpgaaagagactatggpeppdagpa-----995	
QY	810	RVFSLKSHSTVDHVRDQTEESSLPTESEETETNSTLESLSGKKLTQKLKEDSEAS	869
Db	996	-----lydvaeaeerfpaarprspisltishlag-----sagrt	1031
QY	870	TESVPLVCKSASAHNLSSEKKTGHPRTSMLOKSLSVASAKEKTGLAG-----918	
Db	1032	ddapalshtaarssssgslmeigslsvvtrftanitelnsmmistaaapppgtpics	1091
QY	919	-----KTQTAGVETRKQSKPLPKDKETNRNHSNNTETKDPAPQNSNPAEP	967
Db	1092	sylikpeiqlptmtttfae-----iqplpaievt-----ggaqpatgspaqetpagae	1140
QY	968	RKPKSGIMKQQRVNPPTTANSLNPOTTMQKNFDIGEVCPWEVVDLTPGPVPSESK	1024
Db	1141	aapgkpdleelvaltp-----psfrsvdsgsttpspvsesalcipsspk	1187

RESULT 4
ID AAY28563
XX AAY28563 standard; Protein; 1171 AA.
AC AAY28563;

XX	09-NOV-1999	(first entry)	XX
XX			XX
DE		Rat metabotropic glutamate receptor 5 (mGluR5).	XX
XX			XX
KW		Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;	XX
KW		depressant; anaesthesia; phospholipase C; intracellular calcium.	XX
OS		Rattus sp.	XX
XX			XX
PN	WO938975-A2.		XX
XX			XX
PD	05-AUG-1999.		XX
XX			XX
PF	29-JAN-1999;	99WO-US02033.	XX
XX			XX
PR	30-JAN-1998;	98US-0073190.	XX
XX			XX
PA	(UYTE-) UNIV TECHNOLOGY CORP.		XX
XX			XX
PI	Johnson TE, Rikke BA, Sikela JM, Simpson VJ;		XX
XX			XX
DR	WPI: 1999-494093/41.		XX
DR	N-PSDB; AAZ10803.		XX
XX			XX
PT	CNS depressant sensitivity associated a YAC containing murine		XX
PT	tyrosinase and metabotropic glutamate receptor 5 genes, useful for		XX
PT	identifying CNS depressant modulators		XX
XX			XX
PS	Claim 35; Fig 16; 140pp; English.		XX
XX			XX
CC	This sequence is the rat metabotropic glutamate receptor 5 (mGluR5).		XX
CC	This sequence is not directly claimed, but a claimed polypeptide (part		XX
CC	of AAY28562:mouse mGluR5) is specifically not depicted in the rat		XX
CC	mGluR5. mGluR5 are G protein-coupled receptors capable of activating a		XX
CC	variety of second messenger systems. mGluR5 belongs to mGluR Group I. The		XX
CC	binding of an agonist to group I receptors results in the activation of		XX
CC	phospholipase C causing the mobilisation of intracellular calcium. The		XX
CC	mouse mGluR5 sequence correspond to parts of the mouse vrt2		XX
CC	polynucleotide which is associated with central nervous system (CNS)		XX
CC	depressant activity. CNS depressants are used for a variety of purposes		XX
CC	including anaesthesia.		XX
XX			XX
SQ	Sequence 1171 AA;		XX
Quéry Match 3.3%; Score 210; DB 20; Length 1171;			
Best Local Similarity 19.1%; Pred. No. 1.5e-06;			
Matches 179; Conservative 117; Mismatches 397; Indels 244; Gaps 35;			
QY	219	ETWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPPYLECE	273
Db	332	dvkwdyylklrpetnlrnp-----wfgewqhrfqrlegfaqenskyntk-----cn	381
QY	274	NG-----SYKPGWLW-TLSSAIYGL---QPNLVPEFRGV-----MKVDINLQ	311
Db	382	ssltlrthvqskmgfvinaiysmayghnmqmslcpgyagldcmkpidgrkllidslm	441
QY	312	KVDIDQCSDDGWFSGTHKCHLNNSCEMPTKGLGVLGAYECICKAGFYHPGVLVPV-----	366
Db	442	ktnftvgsgdmllfdengdspgyrlylmnfkeng---kdyfdyinvgswdngelkmddev	498
QY	367	-----NNFRRRGPDOHI-SGSTKDVSEE---AYVCLPCREG-----CPFC-----A	403
Db	499	wskknllirsvcepecekgqkvrkgevcswctctpckeneyvfydeytcckacqlgsupt	558
QY	404	DD-SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVYHFRKAKSIRASGLILLE	456
Db	559	ddltgdcdlipqylrvgdpepiaavvfacglilatitfvtfviifdyrdtpvkvssrelcy	618
QY	457	TILFGSELLLYFPVILYFEPSTFRFCILLRWARLLGFATVGTWTLKLRVLKV-----	509
Db	619	illagiclgylctfcilakpkqicylqrligispamsysalvtktnriarilagskkk	678

QY 510 -----FLSRTAQRIPYMTGGRVMRMLAVILLVFWFLIGWTSVCONLEKQISL----- 538
Db 679 ictckprfmsacq-----lviafili-----ciqgiilvalfime 714
QY 559 ---IGOGKTSDDLIFNCLDRWDYMTAVA-BELFLWGLVYLCYAVRTVPSAPHEPRYMA 614
Db 715 pddmhdyrsirevylctntnlgvvtplgnyngllilscfyafrktrnvpantneakyla 774
QY 615 VAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHPLHVTVTIGLLLPK-----PSHS 670
Db 775 fmytctiilwafvpiyf-----gsnykiitmcfsvsatvalgcmfvpkyvillakp 828
QY 671 SNPNRDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEDIRDELKK----- 720
Db 829 ernvrsaftstvtvrmhvgdgkssaaarssslvnlwkrgrssgetlssngksvtwaqne 888
QY 721 -----LYAQLEIYKRKKMITNNPHLOK---KRCSSKGLGRSIMRRITPEIPTSRSQCS 770
889 kstrgqlwrlsvhlnkknprqtavikpfkstenrg-----paaaagggg 936
771 KEDKEGADHTAGTALIRKNPPSSGNTGKSKEETLKNRVFSLKSHSTYDHRDQTEE 830
Db 937 gpgvaga--gnagctatggpeppdagpka-----lydvaea 970
QY 831 SSSLPTESQEEETENSTLESGLKTLQKLKEDSENAESTESVPLVCKSASAINLSSEKK 890
Db 971 eesfpaarprspistlshlag-----sagrtddapsllhsetaarssssgs 1020
QY 891 TGHPRTSMLQKLSVIAASAKETLGLAG-----KTQTAGVEER 928
Db 1021 lmeqisvvtfrtaniselmsmlstaatpgppgtpicssyllipkeiqblptmtfae-- 1078
QY 929 TKSQKPLPKDKEYNRHSNDTETKDPAPONSNAPEE-PRKPKQSGIMKQQRVNPPTAN 987
Db 1079 ---iqplpaievt-----ggaqatgrvpaqetptgaesagpkdleeilvaltppsp---- 1127
QY 988 SLDLPGTQMKDNFDIGEVCPEWYDITLTPGPVPSSEK 1024
Db 1128 -----frdsvdsgsttpnpsvsesalcipsspk 1155
RESULT 5
AAR64253
ID AAR64253 standard; Protein; 1180 AA.
XX
AC AAR64253;
XX
21-JUL-1995 (first entry)
XX Human mGluR5a.
XX DE
XX KW Metabotropic glutamate receptor; mGluR5; mGluR5a; hippocampus;
XX KW cerebellum; pCMV-r7-3.
XX AC
XX OS Homo sapiens.
XX
XX WO9429449-A.
XX PN
XX PD
XX XX 22-DEC-1994.
XX
XX PF 03-JUN-1994; 94WO-US06273.
XX PR
XX PR 04-JUN-1993; 93US-0072574.
XX
XX (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
XX PA
XX PI Daggett L, Ellis SB, Hess SD, Johnson EC, Liaw C;
XX PI Pontsler A;
XX
XX WP1; 1995-036478/05.
XX DR
XX DR N-PSDB; AAQ80419.
XX

PT New DNA encoding human metabotropic glutamate receptor
PT sub:types(s) - and related proteins, probes, RNA, transformed
PT cells and antibodies, useful in treatment, diagnosis and
PT identification of specific modulators
XX
XX
PS Claim 15; Page 90-95; 125pp; English.
XX
CC A human hippocampus cDNA library was screened with DNA encoding
CC the rat mGluR1 receptor. An isolated clone was used to screen
CC human cerebellum cDNA libraries to isolate additional clones. 3
CC 3 splice variants of the human mGluR5 transcript, mGluR5a (given
CC in AAQ80419), mGluR5b (AAQ80520) and mGluR5c (AAQ80421), encoding the
CC proteins given in AAQ80421-55, respectively, were constructed by
CC ligating different overlapping clones. Coding sequences are linked
CC to regulatory elements of pCMV-r7-3 for expression in mammalian
CC cells.
XX
SQ Sequence 1180 AA;
Query Match 3.1%; Score 198; DB 16; Length 1180;
Best Local Similarity 18.9%; Pred. No. 1.2e-05;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;
QY 219 ETEWFHGLRKRWRPHLRHRRGNQGRGLGHSRRKDGGLGDKSHFKWSPPYLECENSGY- 277
Db 333 dvkfwddyyiklrpetnhrp-----wfgewqhr-----fqrleaffqenskyn 378
QY 278 -----KPGMLV-TLSSAIYGL---QPNLVPERGV-----MKVD 307
Db 379 ktcnssltkthvqdkmgfvinaiysmayglnhmqmsicpgyagldcamkpidgrkll 438
QY 308 INLOKVIDIDOCSSDGWFSGTHKCHLNSECMPKGLGLFVLCAYECICAGFYHHPGLVFN 367
Db 439 eslmkntftvgdtilfdengdsprgyelmmfkemg-----kdyfdyinvgsd 488
QY 368 NFRRRGPDQHI-----SGSTKDVSEE-----AYVCLPREG----- 398
Db 489 ngelkmdddewskksnliirsvcepcekgqikvirkgvscwtctpckeneyvfdet 548
QY 399 CPFC-----ADD-SPCFVQEDKYLK-----LAILSFQGLCMLLDVSMVYVHFRKAK 445
Db 549 ckacqlgswptdtdlgtcdilipvqylrwgdpapiaavvfacilgllptlftvrvflliyrdtp 608
QY 446 SIRASGLILLETILFGSLLLYFPVILYFPSTFRICILLRWARLLGFATVGTVTLKLHR 505
Db 609 vvkssrelcyilagicylctfcliaqpkqilycylqriglspamsysalvtktmr 668
QY 506 VLKV-----FLSRTAQRIPYMTGGRVMRMLAVILLVFWFLIGWTSVCON 551
Db 669 iarilagskkkictpkprfmsacq-----lviafili-----ciq 704
QY 552 LERQISL-----IGOGKTSDDLIFNCLDRWDYMTAVA-BELFLWGLVYLCYAVRTV 603
Db 705 lgiivalfimeppdimhdyrsirevylctntnlgvvtplgnyngllilscfyafrktrn 764
QY 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHPLHVTVTIGLL 663
Db 765 panfpeakylaftmytctiilwafvpiyf-----gsnykiitmcfsvsatvalgcmf 818
QY 664 IPK-----FSHSSNNPRDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEDIRD 716
Db 819 vpkvyiilakpernvrsaftstvtvrmhvgdgkssaaarssslvnlwkrgrssgetlss 878
QY 717 ELKK-----LYAQLEIYKRKKMITNNPHLOK---KRCSSKGLGRSIMRRIT 759
Db 759 gngksvswaqneksrgqhlwgrlshlnkknprqtavikpfkstenrglgag----- 932
QY 760 EIPETVSRQSKEDKADHTAGTALIRKNPPSSGNTGKSKEETLKNRVFSLKSHS 819
Db 933 -----agag-gsagvgvatgagcagagpggpespdagppkaly----- 969
QY 820 TYDHRDQTEESSLSLPTESQEEETENSTLESGLKTLQKLKEDSENAESTESVPLVCKS 879

Db 970 -----dvaeeahfpaparprspstishrag-----sastddvpslhsepvars 1019
 QY 880 ASAHNLSSSEKKTGHPR-TSNLQKSLSVIASAKEKTGLAG-----KTQTA 923
 Db 1020 ssgqslmeqissvvtrftaniselmsmlstaapspgvgaaplcssylipkeiqtlptmtt 1079
 QY 924 GVEERTKSKPKPKDKETNRNHSNNTETKDPAPQN--SNPAEPRKPKQKSGIMKQQRV 981
 Db 1080 fae-----lqplalevtggaagaagaagaarrespaagpeaaaakpdeelelvaltp 1134
 QY 982 NPTTANSLNPGTOMKDNFDJGEVCPWEVDYLTGCPVPSESK 1024
 Db 1135 sp-----frdsvdsgsttpnpsvsealcpsspk 1164

RESULT 6

AA49134
 ID AAY49134 standard; Protein; 1397 AA.

XX AAY49134;

07-JAN-2000 (first entry)

DE pmGluR2/Car*Galphai5+3Ala linker fusion construct protein sequence.

XX G-protein fusion receptor; Car; calcium receptor; GluR; head injury;
 KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
 KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
 KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
 KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
 KW cognitive disorder.

XX Homo sapiens.

XX WO951641-AL.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSF-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JB;
 PI Simin RT;

XX WPI; 1999-610995/52.

XX N-PSDB; AAZ31065.

PT New G-protein fusion receptors and chimeras containing domains from
 PT different receptors, used to screen for modulators, potentially useful
 PT e.g. for treating or preventing stroke or Alzheimer's disease

XX Example 1; Fig 12; 255pp; English.

XX The invention relates to G-protein fusion receptors (I) comprising:
 CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
 CC intracellular (ICD) domains, each chosen independently from a Car
 CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
 CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
 CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
 CC linker. (I), and recombinant chimeric receptors (CR) without the GP
 CC component, are used to assess function of the various domains and to
 CC identify compounds (e.g. allosteric modulators or antagonists) that act
 CC on these domains. The modulators are potentially useful for treating or
 CC preventing diseases associated with the receptors, e.g. stroke, head or
 CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
 CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
 CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
 CC (1) for recombinant production of corresponding proteins; and (2) to
 CC produce cells used in screening for modulators. Use of Car and mGluR

CC domains allows presentation of GABABR domains, to a binding agent, in a
 CC form more like the natural domain structure compared with use of
 CC incomplete receptors, lacking one or more domains. By shuffling different
 CC domains, agents can be identified that affect particular domains of a
 CC receptor.

XX Sequence 1397 AA;

Query Match 3.1%; Score 198; DB 20; Length 1397;

Best Local Similarity 21.7%; Pred. No. 1.5e-05;

Matches 131; Conservative 88; Mismatches 236; Indels 148; Gaps 27;

QY 390 YVCLPREGCPF-----CADD-----SPCFVQEDKYLK-----LAIFS 422

Db 520 wlicpcq--pyeyrldefcadcglwywnaslsgcfcelpqeyirwgawavpvtac 576

QY 423 FQGLCMLLDFFVSMVYHFRKAKSTRASGLILLETILFGLSLLLYFPVYILYFEPSTFRCI 482

Db 577 lgalatl--fvlgvvrh--natpvvkasgrelycillggvflcygmtfiflaktstact 633

QY 483 LLRWARLLGFATVYGVTKLHRLVKVF--LSRTAQRIPYMT--GGVVRMLAVI---LL 535

Db 634 lrrlglgtafsvcsailltktnrariifggaregagrprfispasqvaiclaillsgqli 693

QY 536 VVFWFLI---GWTSSVQCNLEKQISLIGQKTSQD---HLIFNMCLIDRWYMTAVAEFLF 589

Db 694 vvawlvveapgtgketaperrevvtlrcnhrdasmglslaynvlli----- 739

QY 590 LLMGVYLC--YAVRT--VPSAFHEPRYMAVAVHNELIIISAIPTIRFVLAS--RLQSDMM 643

Db 740 -----alctlyafktrkpcnfneakfigfmyttciliwlaflpifvvtssdyrvgttm 794

QY 644 LMLYFAHTHTLVTVTIGLLLPKF-----SHSSNNPRDIDATEAYEDELDM----- 689

Db 795 cv-----svslsgsvvlgclfpaklhilifqpknkieevrcstaaahfkvaaratlrnsn 850

QY 690 -----GRSGSYLNSSINSAMSEHSLDPEDIRDELKLYAQLEIYKRRKMIT---N 736

Db 851 vskrssslggstgstpsissksnsedpfqperqkqqlaltqeqqqqpltpqq 910

QY 737 NPHLOKRCRCKK---GLORSIMRRITEPTEVVSROCKEDKEGADHGTA-----K 783

Db 911 qrsqgqprckqkvifgsg-----tvtfslsfdepqknamahgnstqnsleaqs 960

QY 784 GTALTRKNP--PESSGNTGKS---KEETLKNRVFSLKSHSTYDHDVDTQTESSSLPTES 838

Db 961 sdtlrrhqlpllqcgetdldltvqetglqgvvgdqr-----pevedpeelspalvvss 1015

QY 839 QBEETTENSTLESLSGKKLTQKLKEDSEASTESVPLVCKSASAHNLSSEKKTGHPRTSM 898

Db 1016 sq-----sfvisgggstvtenvvnsaaamtlesimacc-----lseeakearrinde 1062

QY 899 LOKSL 903

Db 1063 lerrql 1067

RESULT 7

AA49134
 ID AAY28564 standard; Protein; 1212 AA.

XX AAY28564;

XX 09-NOV-1999 (first entry)

XX Human metabotropic glutamate receptor 5 (mGluR5).

DE Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;
 KW depressant; anaesthesia; phospholipase C; intracellular calcium.

XX Homo sapiens.

Qy	837	ESQSEETENSTLESGLTKLTKL-----KEDSE-----AESTESVPLVCKSASA	888
Db	917	waqnekssr-----gghlwgrlsihinkkenpqntavikpfpkstesrglgagagag	968
Qy	883	HNLSSKKTG-----HPRTSMLOKSLSVIASAKEKTLCLAGKTTQTAGVEERT	929
Db	969	gsagvgatggagcagagpggpespdag--pkalydvaeaeefpapa-----rp	1010
Qy	930	KSQRLPKDKETNRNINSNDTETKDPAPQ--NSNPAEEPRKPKQSGIMKQ--ORVNP	986
Db	1017	rspapi-----stlsrhrgasatddvpsllsepvvarssssqgs--lneqisvvt	1070
Qy	987	N--SDLNPGTTQMKDNDFIGVCPWEYDITPG---PVPSESKVQKHVSIVASEMKN	1042
Db	1071	nisein-----smmlstaap-----spvgaplcssyllipkeiqI-----pt-	1107
Qy	1043	SLREKSHHKPAAEVCCQSNQKRIDKAIEVCLWESQCSILEDEKLILSKTVPVPERA	1107
Db	1108	-----tmttfaeiqpipaiev	1122
Qy	1103	NGG--OPRAANVCAG---QSEELPPKAVASKTENENLNQIGHQEKTKSSSEENV	1155
Db	1124	tgaqaapaagaagdaarespaagpeaaakpdleelval-----tppspfdsvds	1175
Qy	1158	SNFQQLTSGRAEVCPEWETPAQPNAGRSVALPAS	1193
Db	1176	gst-----tpnspvsesalclipss	1194

RESULT 9

AA49129

ID AA49129 standard; Protein; 1394 AA.

AC AA49129;

XX

DT 07-JAN-2000 (first entry)

XX

DE pmluR2/Car*Galphaqi5 fusion construct protein sequence.

XX

KW G-protein fusion receptor; Car; calcium receptor; GluR; head injury; metabotropic glutamate receptor; GABABR; chimeric receptor; stroke; gamma-aminobutyric acid receptor; allosteric modulator; antagonist; spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia; Alzheimer's disease; hyperparathyroidism; osteoporosis; depression; cognitive disorder.

KW

XX Homo sapiens.

OS

XX WO951641-A1.

PN

XX 14-OCT-1999.

PD

XX 02-APR-1999; 99WO-US07333.

PF

XX 03-APR-1998; 98US-0080671.

PR

XX (NPSP-) NPS PHARM INC.

PA

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE; Simin RT;

PI

XX WPI; 1999-610995/52.

DR

XX N-PSDB; AA231060.

DR

XX

XX New G-protein fusion receptors and chimeras containing domains from different receptors, used to screen for modulators, potentially useful e.g. for treating or preventing stroke or Alzheimer's disease

PT

XX Example 1; Fig 12; 255pp; English.

PS

XX

CC The invention relates to G-protein fusion receptors (I) comprising: (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) intracellular (ICD) domains, each chosen independently from a Car

QY 709 LDPEDIRDELKLYLAQLEYKRRKMTNNPHLOKKRCS-----KKGLGRSIMRRITEI 761
Db 875 -----ifrrkkaagnansngksvsepgggvqgqhmwhrlsvh 916
QY 762 PETVSRQCKEDKADHGTAKTALIRKNPPSSNGTCKSEETLKNRVSLKKSHSTY 821
Db 917 vknetaacq-----tavik-----pltksyggsgksltf----- 946
QY 822 DHVRDQTESSSLPTSEBETNTSTLESGLKTLQKLKDESEASTESVPLVCKSAS 881
Db 947 -----sdtst-----klynveeedaqpirfpgpspsmv 977
QY 882 AHNLSSEKKTGHPRTSMLOKSLSVIAASAKETLGLAGKTQTACGVBERTKSQPLPKDKET 941
Db 978 vhrvpsaatpplppl-----taetpfla-----epalpg----- 1012
QY 942 NRHNSDNTETKDPAPQNSPAEPRKQKSGIMKQOR---VNPTANSDLN----- 991
Db 1013 -----lppliqggqg-----ppqksimdlqldgsvsnfstai.pdfhavlagpg 1056
992 -PGTQMKDNFDIGVCPWEYDITPGVPSE--SKVQKHVSIVASEMEKNPTFSLKRS 1048
1057 gpgn-----glrslpyppppqhlqmplqlstfgeelvppadddsdse 1101
QY 1049 HHKPKAAEVCOQSNQKRIDKAEVCLWESOGSILEDEKLLISKTPVLPRAKEENGQPR 1108
Db 1102 rfklldqeyvveheregteede-----leeeedlgaaskltpddspaltppspf 1151
QY 1109 AANVCAGQSEELPP--KAVASKTENENLQIGHQEKKTSS 1147
Db 1152 rdevagssvpsspsvsvictppnvsyavilrdykqss 1192

RESULT 11
AAW25765
ID AAW25765 standard; Protein; 1056 AA.
XX
AC AAW25765;
XX
DT 03-DEC-1997 (first entry)
DE Amino acid sequence of pCH4.
XX
KW Chimeric receptor; extracellular domain; seven transmembrane domain;
KW intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
KW MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;
KW neurological disease.
XX
KW Chimeric - Homo sapiens.

Key Location/Qualifiers
FT Misc-difference 137
FT /note= "Encoded by MTS"
FT Misc-difference 146
FT /note= "Encoded by WCC"
XX
XX WO9705252-A2.
XX
XX 13-FEB-1997.
XX
XX 25-JUL-1996; 96WO-US12336.
XX
XX 26-JUL-1995; 95US-0001526.
XX
XX (NPSF-) NPS PHARM INC.
XX
XX Fuller FH, Hammerland LG, Krapcho KJ;
XX WPI; 1997-145690/13.
XX N-PSDB; AAT86168.
XX
XX Chimeric receptors comprising metabotropic glutamate receptor and
XX calcium receptor - used for screening for neurologically active

compounds

Example 5; Fig 5; 177pp; English.

This sequence represents a chimeric receptor. The chimeric receptor comprises an extracellular domain, a seven transmembrane domain, and an intracellular cytoplasmic tail domain, and a sequence of at least 6 contiguous amino acids is homologous to a sequence of a metabotropic glutamate receptor (MGR), and a sequence of at least 6 contiguous amino acids is homologous to a sequence of a calcium receptor (CR). The chimeric receptor may be used for screening for compounds that bind to or modulate the activity of MGR or CR and for determining the site-of-action of a CR active compound. The compounds can be used in the treatment of neurological diseases and disorders. They can also be used as diagnostic agents. Chimeric receptors such as this, allow the coupling of certain functional aspects of an MGR with certain functional aspects of a CR. They allow for more efficient high-throughput screening of compounds.

XX Sequence 1056 AA;

Query Match 3.0%; Score 190; DB 18; Length 1056;
Best Local Similarity 19.9%; Pred. No. 3.9e-05;
Matches 124; Conservative 97; Mismatches 241; Indels 162; Gaps 20;

QY 357 GFYHPGVLPVNNFRRRGPDQHSSTKDYSE-----AYVCLPCREG- 398
Db 498 gtwhgvinidyykq---mnksgvrvscvsepclgkqikvirkgevsccwictackene 554
QY 399 -----CPFC-----ADSPCFVQEDKYL-----LAIISFGQLCLMDFVSMLVV 438
Db 555 yvqdefctkacdlgwwpnadltgcepiprvylewsniepilaiafscgilvltflif 614
QY 439 YHFRKAKSIRASGLILLETILFGLSLLLPVPPVILFEFTFCILRLNARLGFATVGT 498
Db 615 vlyrdtpvkvssrelcyllagiflgvvcptllakpttscylqrlilvgissamcysa 674
QY 499 VTLKLHRLVKV-----FLSRTAQRIPYMTGGRVMRLAVILLVVFVFLIGW 544
Db 675 lvtktoriarilagskkcltrkprfmsawaqviailisvqltlvtllii----- 726
QY 545 TSSVCONLEKQISLIGQKTSDDLIFNMCLIDRWDMYMTAAEF-----LFLMGVYLCYAV 600
Db 727 -----meppmpil-----syysikevlyicntsnlgvavplgynllimsctyyafkt 774
QY 601 RTVPSAFHEPRYMAVAVHNELIISAIFHTIRFVLSRQLQSDHMLMYFAHTHTLTVTVIG 660
Db 775 rnvpanfneakyiaftmyttclilwafvplyf-----gsnykiittcfavslsvtvalg 828
QY 661 LLLIPK-----FSHSSNNPRDDI--ATEAYEDELD-----GRSGSYL 696
Db 829 cmftpkmyiliakperntieevrcstaahafkvaaratlrnsnvarkrssslggsgtsp 888
QY 697 NSSINSAMSEHSLDPEDIRDELKLYLAQLEYKRRKMT---NNPHLOKKRCSK---GL 750
Db 989 sssissksnsedpfqperqkqqlalttqqeqqqpltpqqqrsggprckqkvifs 948
QY 751 GRSIMRRITEIPETVSRQCKEDKADHGTAKTALIRKNPPSSNGTCKSEETLKNR 810
Db 949 g-----tvtfslsfdepqknmahgns-----thqns 975
QY 811 VFLKKSHSTYDHVRDQTESSSLPTSEBEE---TTENSTLESGLKTLQKLKDESEA 867
Db 976 leaqksdtl-----trhaplplqcggetdldltvqetlgqgpvgdqrpevedpeel 1028
QY 868 ESTESVPLVCKKSASAHNLSSEKKT 891
Db 1029 ----spalvvsssqsfvisgggst 1048

RESULT 12
AAW25764

CC intracellular (ICD) domains, each chosen independently from a Car
 CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
 CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
 CC linker. (1), and recombinant chimeric receptors (CR) without the GP
 CC component, are used to assess function of the various domains and to
 CC identify compounds (e.g. allosteric modulators or antagonists) that act
 CC on these domains. The modulators are potentially useful for treating or
 CC preventing diseases associated with the receptors, e.g. stroke, head or
 CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
 CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
 CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
 CC (1) for recombinant production of corresponding proteins; and (2) to
 CC produce cells used in screening for modulators. Use of Car and mGluR
 CC domains allows presentation of GABAR domains, to a binding agent, in a
 CC form more like the natural domain structure compared with use of
 CC incomplete receptors, lacking one or more domains. By shuffling different
 CC domains, agents can be identified that affect particular domains of a
 CC receptor.

XX Sequence 1276 AA;

Query Match 2.9%; Score 183.5; DB 20; Length 1276;

Best Local Similarity 21.8%; Pred. No. 0.00016;

Matches 111; Conservative 74; Mismatches 203; Indels 121; Gaps 24;

QY 311 QKVDIDQSSDGSFSGTHKC---HLNNSC-CMPDKGLGFLVAYECICRAG---FYHGV 363

DB 475 eqvtfdcc--gdlvgnyslinwhlspedgsivfkevgv---ynvyakkgerlfinck 527

QY 364 LPVNNFRRGPDQH-----ISGSKTDVSEE---AYVCLPCRGE-----CPFCAD 404

DB 528 ilwsgtsrevfscnrcdclagtrkgliegeptccfecvcpdgeysdetdasackncpd 587

QY 405 D-----SPCFVQEDKYLR-----LAISFGCLMLDLDFVSMVYHFRKAKSIR 448

DB 588 dfwsnennhtscfelpqeyirwdgawavpvtiaclgalatl--fvlgvfrh-natpvvk 644

QY 449 ASGLILETILFGLSLYFPVYLYPEPTFCILLRWALLGFATVYGTITLKLHRLVK 508

DB 645 asgrelyilggvfyicymtftiakpbtavctlrilglgtafsvcsalltktnr 704

QY 509 VF--LSRTAQIPYMT--GGRVMMLAVI---LLVVFWEFI---GWTSSVQNLEKQISL 558

DB 705 ifggarregagrprfispasqvalclalissgqllivvawlvveapgtgketape 764

QY 559 ICGGKTSQD---HLIFNMCLIDRWYMTVAEFLFLMGVYLCYAVTVPSPAFHEPRYMAV 615

DB 765 rchrdasmllgslyaynvlil-----aictly-afntrkcpenfneakfigf 809

QY 616 AVHNELIISAIFHTIRFVLAS--RLQSDWMLMLYFAHTHTLTVTITGLLIPKF----- 667

DB 810 tmyttciilwllalpiyvressyrvqttmcv-----svslsgsvlglcfapklhllfq 865

QY 668 -----SHSNPNRDIATAYEDELDMGRSGSY-----LNSSIN 701

DB 866 pqknvshraptrsfgsaaarssslqgsgsfvptvcngrevvdstsslmleslma 925

QY 702 SAWSEHSLOPEDIRDELKLYAQLEIYKR 730

DB 926 ccalseakearrindelefer---qlrrdr 951

RESULT 14

AAR25081

ID AAR25081 standard; Protein; 906 AA.

XX

AC AAR25081;

XX

DT 04-JAN-1993 (first entry)

XX

DE GLU-G-R subtype 1b.

XX GLU-G-R; Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;
 KW neurotransmitter; glutamate; neuronal excitation.

XX Synthetic.

XX WO9210583-A.

XX 25-JUN-1992.

XX 12-DEC-1991; 91WO-US09422.

XX 12-DEC-1990; 90US-0628806.

XX 30-JAN-1991; 91US-0648481.

XX 18-MAR-1991; 91US-0672007.

XX (UNIW) UNIV WASHINGTON.

XX (ZYMO) ZYMOGENETICS INC.

XX Almers W, Hagen FS, Houamed KM, Mulvihill ER;

XX WPI: 1992-234638/28.

XX N-PSDB; A925815.

XX Mammalian G protein-coupled glutamate receptors - activate
 PT phospholipase C or stimulate inositol phospholipid metabolism,
 PT for use in diagnosis and identification of receptor agonists and
 PT antagonists

XX Disclosure; Fig 7; 144pp; English.

XX The sequence given is mammalian G protein-coupled glutamate
 CC receptor (GLU-G-R) subtype 1b. The clone encoding this protein was
 CC isolated from Xenopus oocyte cultures which had been injected with RNA
 CC extracted from a cerebellum sublibrary. Expression of GLU-G-R
 CC activity was assessed by voltage clamp assay. Plasmid DNA encoding
 CC GLU-G-R activity was isolated by replica-plating experiments. The
 CC protein encoded by clone 45-A was found to act as a receptor for the
 CC neurotransmitter glutamate, and may play a critical role in the control
 CC of neurons. GLU-G-R and antibodies may be used to identify agonists
 CC and antagonists of GLU-G-R-mediated neuronal excitation and in
 CC diagnosis.

XX Sequence 906 AA;

Query Match 2.8%; Score 180; DB 13; Length 906;

Best Local Similarity 20.6%; Pred. No. 0.00017;

Matches 93; Conservative 66; Mismatches 170; Indels 122; Gaps 15;

QY 357 GFYHPGVLPVNNFRRRGPDQHSSTKDVSEE-----AYVCLPCRGE- 398

DB 498 glwhegvlnidkykq---mnksgmrvscvsepclkgkivirkgevsccwictackene 554

QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAISFGCLMLDLDFVSMV 438

DB 555 fvqdeftrcradlgwvphaelgtcepiprvylewsdiesilaafscgilvltvlif 614

QY 439 YHFRKAKSTRASGLILLETILFGLSLLYFPVYLYPEPTFCILLRWALLGFATVYGT 498

DB 615 vlyrdtpvkvssrelcyllagiflygvcpftliakpttscylqrlivgissamcysa 674

QY 499 VTLKLHRLVKV-----FLSRTAQIRIPVMTGGRVMRLAVTLVVFVFLIGW 544

DB 675 lvtktnrliarilagskkkictrkprfmsawaqviiasllisvqltlvltlii----- 726

QY 545 TSSVQNLEKQISLIGQGTQSDHLIFNMCLIDRWYMTVAE-EFLFLMGVYLCYAVRTV 603

DB 727 -----mepmpilyspsike--vyllicntslngvavpvgngllmsectyyafktrv 777

QY 604 PSAFHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHTLTVTITGLL 663

DB 778 panfneakyiaftmyttciilwlaflayf-----gsnykiittcfavslsvlalgcmf 831

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 08:59:27 ; Search time 26.44 Seconds
(without alignments)
946.190 Million cell updates/sec

Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGAMAYPLLLLLAQLGLG.....LSANKIAGPRKEIWDSPKV 1215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	220.5	3.5	1199	1	US-08-041-538-2
2	220.5	3.5	1199	1	US-08-463-642-2
3	220.5	3.5	1199	1	US-08-455-802-2
4	220.5	3.5	1199	2	US-08-465-157-2
5	220.5	3.5	1199	5	PCT-US91-09422-2
6	220	3.4	1219	2	US-08-687-289A-6
7	204	3.2	1180	4	US-08-660-148-2
8	200	3.1	1180	3	US-08-486-270-8
9	200	3.1	1180	3	US-08-367-264-8
10	197.5	3.1	1212	4	US-08-660-148-5
11	193.5	3.0	1212	3	US-08-486-270-10
12	193.5	3.0	1212	3	US-08-367-264-10
13	193	3.0	1180	3	US-08-072-574-8
14	190	3.0	1056	2	US-08-687-289A-8
15	186.5	2.9	1212	1	US-08-072-574-10
16	186	2.9	1056	2	US-08-687-289A-7
17	178	2.8	571	4	US-08-961-083-4
18	177	2.8	906	1	US-08-486-270-2
19	177	2.8	906	3	US-08-367-264-2
20	174	2.7	906	5	PCT-US91-09422-17
21	173.5	2.7	1964	2	US-08-790-912-3
22	173.5	2.7	2052	2	US-08-790-912-2
23	172.5	2.7	1079	1	US-08-485-588-8
24	172.5	2.7	1079	1	US-08-484-565-8
25	172.5	2.7	1079	2	US-08-480-751-8
26	172.5	2.7	1079	2	US-08-943-986-8
27	172.5	2.7	1079	3	US-08-353-784-8

Query Match 3.5% Score 220.5; DB 1; Length 1199;

ALIGNMENTS

RESULT 1
US-08-041-538-2
; Sequence 2, Application US/08041538
; Patent No. 5385831
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA: US/08/041,538
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1199 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-041-538-2

Sequence 8, Appli
Sequence 12, Appl
Sequence 12, Appl
Sequence 6, Appli
Sequence 6, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 12, Appl
Sequence 12, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 26, Appli

```
Best Local Similarity 18.6%; Pred. No. 2.1e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

QY 357 GFYHGVLPVNNRRRGGPDQHSSTKDVSEE-----AYVCLPCREG- 398
Db 498 GTWHEGVNLIDDKYIQ---MNSGMVRSVCSEPCCLKGQIKVIRKGEVSCCWTACKENE 554
QY 399 -----CPFC-----ADDSPCFVQEDKYLK-----LAISFQGLCMLLDFVSMVLV 438
Db 555 FVODEFTCRACDLGWWPNAELTGCCEPIPVRYLEWSDIESIIAIAFSCGLILVTLFVLIF 614
QY 439 YHFRKAKSTRASGLILLETILFGLSLLYFPVILYFEPSTFCILLRWALLGFATVYGT 498
Db 615 VLYRDPVVKSSRELXYILLAGIFLGYVCPPTLAKPTTSCYQLRLLVGLSSAMCYS 674
QY 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGRVWRMLAVILLVVFVFLIGW 544
Db 675 LVTKTNRIARILAGSKKIKTRKPRFMSAWAQVIAIISVQLTLVVTLLI----- 726
QY 545 TSSVCONLEKQISLIGOGKTSDFHIFNMCLIDRWDMYMTAVA-EFLFLMGVYLCYAVRTV 603
Db 727 -----MEPPWPILSYPSIKE--VILICNTSNLGVAPVGYNGLLIMSCITYAETKRN 777
QY 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMIFYAHTHLTVTVTIGLL 663
Db 778 PANFNEAKYIAFTMTYTCIWLAFVPIVF-----GSNKKIITTCFAVSLSVTVVALGCMF 831
QY 664 IPK-----FSHSSNPRDDIATEAYEDELDMG-----RSGSYLNSSINSANSEHSLDP 711
Db 832 TPKMYIIIAKPERNVRSFTT---SDVVRMHVGDGKLPCRSTNFL----- 874

QY 712 EDIRDELKLYAQLETKYKRRKMTNPNHLOKRC-----KGLGRSIRMRITRIPET 764
Db 875 -----IFRKKKPCAGNANSNGSVSWSEPGGROAPKGOHVQWRLSVHVKT 919
QY 765 VSRQSKDEKAGDHCTAGTALIRKNPSSGNGTKSKEETLKNRVFSLKKSHTYDHW 824
Db 920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948
QY 825 RQUTRESSLPTSEBETSTLESGLKLTOKLKEDSEASTESVPLVCKSASAHN 884
Db 949 -----ASTLYNVEEDNTPSAHFSPSPSPSMVVR 980
QY 885 LSSEKKTGHPTSMLOKLSVSIASAKETLGLAGKTQTAGVBEERTKSKQPLPKRETNRN 944
Db 981 RGPVATTPPLPHL-----TAEETPLFADSVIPKGL-----PPLPQQO----- 1021
QY 945 HNSDNTETKDPAPQNSPAEPRKPKSGIMKQQRVNPTTANSDL-----NPGTQ 996
Db 1022 -----PQPPPPQPPQPPK--SLMDQLQGVVTFNFGSGIPDFHVLGPGTP- 1065
QY 997 MKDNFDIGEVCPWEVYDLPGPVPSE--SKVKQHVSIIVASEMEKNPTFSLKESHHKPKA 1054
Db 1066 -----GN-----SURLVPPPPPHQLQMLPLHSTFQESISPPGEDIDDD 1108
QY 1055 AEVCOOSNOKRIDKAEVCLWESQG---QSILEDKILISKTPVLPRAKEENGOPRAAN 1111
Db 1109 -----ERKLLAQEFVYREGTEDELEEDLPTASKLTPEDSPALTPSPFRDS 1159
QY 1112 VCAGOSEELPP--KAVASKTENENLNQHGQEKYSSS 1147
Db 1160 VASGSVSPSPVSESVCLTPPNVTYASVILRDYKQSSS 1197

RESULT 2
US-08-463-642-2
; Sequence 2, Application US/08463642
; Patent No. 5721107
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
```


Db	1160	VAGSSVPSSPVSESVLCTPPNVTYASVILRDYKQSSS	1197
----	------	---------------------------------------	------

RESULTS

```

US-08-465-157-2
; Sequence 2, Application US/08465157
; Patent No. 5869609
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Hougmed, Khaled M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1  FILING DATE:
2  CLASSIFICATION:
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER:  US/08/041,538
5  FILING DATE:
6  APPLICATION NUMBER:  US/07/648,481
7  FILING DATE:
8  APPLICATION NUMBER:  US 07/626,806
9  FILING DATE:  12-DEC-1990
10 ATTORNEY/AGENT INFORMATION:
11 NAME:  Parmelee, Steven W
12 REGISTRATION NUMBER:  31,990
13 REFERENCE/DOCKET NUMBER:  13952-6-
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE:  206-467-9600
16 TELEFAX:  206-623-6793
17 INFORMATION FOR SEQ ID NO:  2:
18 SEQUENCE CHARACTERISTICS:
19     LENGTH:  1199 amino acids
20     TYPE:  amino acid
21     TOPOLOGY:  linear
22     MOLECULE TYPE:  protein

```

Query Match 3.5%; Score 220.5; DB 2; Length 1199;
Best Local Similarity 18.6%; Pred. No. 2.1e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

[illegible]

675	LVTKTNRIRIAGSKKICTRKPRPMSAWAQVVIISILISVOLTLVTVLLII-----	726
QY	345 TSSVCONLEKQISLIGOGKTSDLHIFNMLCLDRWDYMTAVA-EFLFLMGVYLVCYAVRTV	603
Db	727 -----MEPPMPLSPYSIKE--VYLICNTSLNGWAPGVGNGLLIMSCSYIAFKTRNV	777
QY	604 PSAPFHPRYWAVAVHNELLISAFHTIRFVSLASRLQSDWMLMFAHTHLTVTVTIGLLL	663
Db	778 PANENEAKYIAFTMYTCTIWLAFPIYF-----GSNYKIITCFVAVSLSVTVALGCMF	831
QY	664 IPK-----FSHSSNNPRDDIATEAYEDELDMG-----RSGVYLSNINSINAWSEHS	711
Db	832 TPKWYIIIAKPERNVSFAETT--SDVVRMHVCDCKLPCRNTFNL-----	874
QY	712 EDIRDELKLYIAOLEIYKRRKMITNPHLQKRC-----KKGLGRSIMRRITEIPET	764
Db	875 -----IFRKKPKGAGNANSNGKSVSWSPGGROAPKGOHVWQRLSVHVKT	919
QY	765 VSRQCKSEKEDGADHCTAGTALIRKNPESSGNTGKSKEETLNKRVFSLKKSHTYDHV	824
Db	920 NETACMQ-----TAVIK--PLTKSYOGSGKSLTFSD-----	948
QY	825 RDOTESSSLPTESQBEETENSTLESJGKKLTOKLDESEASTESVPLVCKSAAHN	884
Db	949 -----ASTKTLNVEEDNTPTSAHFSPSPSPSMVVHR	980
QY	885 LSSEKKTGHPRTSMLOKSLSVLASAKETLGLAGTQTQGYVEERTKSKPLPKDKETNRN	944
Db	981 RGPVATVTPPLPPLH-----TAEETPLFVADSVIPKGL-----PPLPQQO-----	1021
QY	945 HNSDNTETKDPAPQNSNPAEPRPKQSGIMKQORVAPTTANSDL-----NPGTTQ	996
Db	1022 -----PQOPPPQPPQPPK--SLMDQLQGVVTFNFGSIGPDPFHAVLAGPCTP-	1065
QY	997 MKDNFDIGVCWPWEYDLTPGPVPSE--SKVQKHVSIVASEMENKPTSLAEKSHHPKA	1054
Db	1066 -----GN-----SURLYPPPPPHQLQMLPLHLSTFQESISPPGGBIDDDDS-----	1108
QY	1055 AEVCQSQNOKRIDKAEVCLWESOG--OSILEDKLLISKTPVLPRAKEENGGOOPRAAN	1111
Db	1109 -----ERFKLLOEFVYEREGNTEDELEEDLPTASKLTPEDSPALTPPSPRDS	1159
QY	1112 VCAQOSEELPP--KAVASKTENENLQIGHOEKTTSS	1147
Db	1160 VASGSSVPSPPSVSESLCTPPNVTYASVILRDYKOSS	1197

```

RESULT - 5
PCT-US91-09422-2
; Sequence 2, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435

```

;; PRIOR APPLICATION DATA: US 07/672,007
;; FILING DATE: 18-MAR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/648,481
;; FILING DATE: 30-JAN-1991
;; PRIOR APPLICATION DATA: US 07/626,806
;; FILING DATE: 12-DEC-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Parmelee, Steven W.
;; REGISTRATION NUMBER: 31,990
;; REFERENCE/DOCKET NUMBER: 13952-6PC
;; TELEPHONE: (206) 467-9600
;; TELEFAX: (415) 543-5043
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1199 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US91-09422-2

Query Match 3.5%; Score 220.5; DB 5; Length 1199;
Best Local Similarity 18.6%; Pred. No. 2.1e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

QY 357 GFYHPGLVPLNFRPRPDQHSSTKDVSEE-----AYVCLPCREG-398
Db 498 GTWHEGVNLDDYKIQ---MNKSMVRSCVSEPCLKQGIKVRGEVSCCMCTACKENE 554
QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAITSFQGLCLMLDFVSLVV 438
Db 555 FVQDEFTCRACDLGWNALTECEPTVRYLEWSDIESIIATFSCGLILVTLVFLIF 614
QY 439 YHPRKASIRASGLILETLFGSLLYFPVILYFPSTFRCLLRLKARLLGPAVYGT 498
Db 615 VLYRDTPWAKSSRELCCYIILAGIFLGVCVPTLIARPTTSCYLRLLVGLSSAMCYSA 674
QY 499 VTLLKLRVLKLV-----FLSRTAQRIPTYMTGGVRMRLAVILLVFWFVLIGW 544
Db 675 LVYKTNRIARILAGSKKIKTRPRFMSAWAQVFIASILISVQLTVTLVTLII----- 726
QY 545 TSSVQCNLEKQISLGQKTSDDLIFNMCLTDRWDMYTA-EPFLFLWGVLYCYAVRTV 603
727 -----MEPPMPLSYPSIKE--VYLICNTSLGVVAPVCYNGCLLMSCTYYAFKTRNV 777
QY 604 PSAPHEPRYAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHTVTVTIGILL 663
778 PANFNEAKYIAFTYTTCTIILAFVPIYF-----GSNYKIITTCFAVSLSVTVLALGCMF 831
QY 664 IPK----FSSHNPNRDIATEAVEDELDMG-----RSGSVLNSINSANSEHSLDP 711
832 TPKMYIIIAKPERNVSFAFT---SDVVRMHVGDGKLPCRSNTFLN----- 874
QY 712 EDIRDELKLYAOEIKYKRMKMTNPNHLOKRCs-----KKGLGRSMRRITETIPET 764
875 -----IFERRKPGAGNANSNCKSVSWSEPGRQAPKQHVWVORLSVHVKT 919
QY 765 VSRQCKEDKAGDHGTAKTALIRKKNPSSNGTGSKEETLKNRVFLSKKSHSYDHY 824
920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948
QY 825 RDQTESSSLPTESQEBETTENSTLESGLKLTQKLKEDSEABESTESVPLVCKSAHN 884
949 -----ASTKTLYNVEEDNTPSAHFSPSPSPSVVHR 980
QY 885 LSSEKKTGHPRTSMLOKSLVSIASAKETLGLAGKTQTAGVEERTKSOKPLPKDKETNRN 944
981 RGPVATTPPLPHL-----TAETPLFLADSVIPKGL-----PPPLPQQQ----- 1021

QY 945 HSNSDNTETKDPQNSNPNAEPRKPKSGIMKQORVNPPTTANSDL-----NPGTTQ 996
Db 1022 -----PQPPPPQPPQPPK--SLMDLQGVVTFNGSGIPDFHFAVLAPGPTP- 1065
QY 997 MKDNFDIGEVCPWEVYDLTPGPVPE--SKVQKHVSIVASEMEKNPTFSLKEKSHHKPKA 1054
Db 1066 -----GN-----SLRSLYPPPPPPQHLQMLPLHLSTFQESISPPGEDIDDD- 1108
QY 1055 AEVCQSQNQRIDKAEVCLMESQG---QSILEDKLLISKTPVLPRAKEENGQOPRAAN 1111
Db 1109 -----ERFKLLQEFVYVEREGTEDELEBEDLPTASKLTPEDSPALTPSPFRDS 1159
QY 1112 VCAGQSEELPP--KAVASKTENENLNOIGHQEKTTSS 1147
Db 1160 VAGSSVPSFVSESVLCTPPNVTYASVILRDYKQSS 1197

RESULT 6
US-08-687-289A-6
; Sequence 6, Application US/08687289A
; Patent No. 5981195
; GENERAL INFORMATION:
; APPLICANT: Fuller, Forrest H.
; APPLICANT: Krapcho, Karen J.
; APPLICANT: Hamerland, Lance G.
; TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
; IDENTIFYING COMPOUNDS ACTIVE AT
; TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
; TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
; TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS
; TITLE OF INVENTION: AND DISEASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Fastseq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: July 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-687-289A-6

Query Match 3.4%; Score 220; DB 2; Length 1219;
Best Local Similarity 18.8%; Pred. No. 2.3e-09;


```
QY 556 ISL-----IGOKTSDHLIFNMCLIDRDWYMTAVA-BELFLLGWYLCYAVRTVPSAF 607
Db 709 VALFIMEPPDMDHYPSIREVYLCTITNLGVVTPLGYINGLLILSCTFYAFKTRNVPANF 768
QY 608 HEPYMAVAVHNELIIISAIFHTIRFVLASRLQSDWMLMIFYAHTHLTVTTIGLLIPK- 666
Db 769 NEAKYIAFTMYTTCIIWLAFVPIYF-----GSNYKIITMCFVSLSATVALGCMFVPKV 822
QY 667 ---FSHSSNNPRDDIATE-AYEDELMDGRSGSYL--NSSINSAMSEHSLEDIPDIRDELK 720
Db 823 YIILAKPERNVRSAFTTSTVVRMHVGDGKSSAASRSSLVNLWKRRGSSGETLSSNGKS 882
QY 721 -----LYAQLEIYKRRKMITNNPHLOK---KRCSSKGLGRSIRMRIPE 763
Db 883 VTWAQNEKSRGQHLWQRLSIHINKENPNQTAIVKPPKSTESRGILAG----- 932
QY 764 TVSRQSKEDKEGADHGTAKGTALIRKNPPESGNTGKSKEETLKNRVFSLKKSHSTYDH 823
Db 933 -----AGAG-GSAGGVGATGAGCAGAGCGPESPDAGPKALY----- 969
QY 824 VROTESSSLPTESQEEETTENTSTLESGLKKLTOKLKEDSEAEATESVPLVCKSASAH 883
Db 970 --DVAEAEHFAPARPRSPSTLSHRAG---SASRTDDDDVPSLHSEPVARSSSOG 1023
QY 884 NLSEKKTGHPR-TSMLQKLSVIAASAKEKTLGLAG-----KTQTAGVEE 927
Db 1024 SLMEQISSVVYTRFTANISELNSMMLSTAAPSPGVGAPLCSYLIPIKEIQLPTMTTFAE- 1082
QY 928 RTKSQKPLPKDKETNRNHSNDTETKDPAPON--SNPAEPRKPKQSGIMKQORVNPTT 985
Db 1083 ----IQPLPAIEVTGGAQPAAGAAQADARESAPAGPEAAAKPDLEELVALTPPSP-- 1136
QY 986 ANSDLNPGTTQMKDNFDIGEVCPWEVYDLTPGPVPSESK 1024
Db 1137 -----FRDSVDGSTTPNSPVSESALCIPSSPK 1164

RESULT 8
US-08-486-270-8
; Sequence 8, Application US/08486270
; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
```

```
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-486-270-8

Query Match 3.1%; Score 200; DB 1; Length 1180;
Best Local Similarity 18.9%; Pred. No. 11e-07;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;

QY 219 ETWFHCLRLKRWPHLRGPNOPRGLGHSWRKDGGLGDKSHFKWSPPLYCECNSY- 277
Db 333 DVKWFDDYILKLRPETNHRNP-----WFQEFWQHR-----FOCRLEAFPOENSKYN 378
QY 278 -----KPGWLVTLSAIYGL---QPNLVPEPRGV-----MKVD 307
Db 379 KTCNSSLTATHRVQDSKMGFVINAISYMAVGLHNMQMSLCPGAYAGLCDAMKPIDGRKLL 438
QY 308 INLQKVIDQCSSDGMFSGTHKCHLNNSCMPKGLGFLVGLGAYECICKAGFYHGPVLVN 367
Db 439 ESLMKTNFTGSGDTILFDENGSDPGRYEIMNFKMG-----KDYFDYINVGSWD 488
QY 368 NFRRRGPDQHI-----SGSTKDVSEE-----AYVCLPREG----- 398
Db 489 NGELKMDDEWVSKSNIIRSVCSPECEKQIKVIRKEVSCCWTCTPCKENEYVFDEYT 548
QY 399 CPFC-----ADD-SPCFVQEDKYL-----LAIISFOGLCMLDFFVSLVYVHERKAK 445
Db 549 CKACQLGSWFTDGLTGCDLIPVQVLRWGDPETAAVFAACLGALLATLVTVVFIYRDTP 608
QY 446 SIRASGLILLETILFGSLLLYFPVVIYLFEPSTFRCLLRWARLLGPATVYGVTVTKLHR 505
Db 609 VKSSSRRELCYIILAGICLGYLCTFCLIAKPKQIYCYQLRIGIGLSPAMSYSALVTKTNR 668
QY 506 VLKV-----FLSRTAQRIPYMTGGVRMRLAVILLVWFVFLIGHTVSSVCQN 551
Db 669 IARILAGSKKIKCTPKPRFNSACAQ-----LVIAFILI-----CIQ 704
QY 552 LEQOISL-----IGOKTSDHLIFNMCLIDRDWYMTAVA-BELFLLGWYLCYAVRTV 603
Db 705 LGIIVALFIMEPPDMDHYPSIREVYLCTITNLGVVTPLGNNGLLILSCTFYAFKTRNV 764
QY 604 PSAPHEPRYMAVAVHNELIIISAIFHTIRFVLASRLQSDWMLMIFYAHTHLTVTTIGLL 663
Db 765 PANFPEAKYIAFTMYTTCIIWLAFVPIYF-----GSNYKIITMCFVSLSATVALGCMF 818
QY 564 IPK-----FSHSSNNPRDDIATE-AYEDELMDGRSGSYL--NSSINSAMSEHSLEDIPDIR 716
Db 819 VPKYIILAKPERNVRSAFTTSTVVRMHVGDGKSSAASRSSLVNLWKRRGSSGETLSS 878
QY 717 ELKK-----LYAQLEIYKRRKMITNNPHLOK---KRCSSKGLGRSIRMRI 759
Db 879 NGKSVTWAQNEKSRGQHLWQRLSIHINKENPNQTAIVKPPKSTESRGILAG----- 932
QY 760 EIPETVSRQSKEDKEGADHGTAKGTALIRKNPPESGNTGKSKEETLKNRVFSLKKSHS 819
Db 933 -----AGAG-GSAGGVGATGAGCAGAGCGPESPDAGPKALY----- 969
QY 924 TYDHRDQTESSSLPTESQEEETTENTSTLESGLKKLTOKLKEDSEAEATESVPLVCKS 879
Db 970 -----DVAEAEHFAPARPRSPSTLSHRAG---SASRTDDDDVPSLHSEPVARS 1019
QY 880 ASAHNLSSEKKTGHPR-TSMLQKLSVIAASAKEKTLGLAG-----KTQTAG 923
Db 1020 SSQGLMEQISSVVYTRFTANISELNSMMLSTAAPSPGVGAPLCSYLIPIKEIQLPTMTT 1079
```

[illegible]

RESULT 9
US-08-367-264-8
; Sequence 8, Application US/08367264
; Patent No. 6001581
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Prety, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071

```

Query Match          3.1%; Score 200; DB 3; Length 1180;
Best Local Similarity 18.9%; Pred. No. 1.1e-07;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;

QY 219 ETWFHCLRKWKPHLHRRCPNOCPRCLGHSWRKDCGLGDGKSHFKWSPPYLECEGNSY- 277
   : : | | | | | | | | | | : : : : | | |
Db 333 DVKWFDDYYLKLRPETNHRNP-----WFOEFWOHR-----FQCRLEAFPOENSKYN 378

QY 278 -----KPGWLW-TLSSATYGL---QPNLVPFRGV-----MKVD 307
   : : : : : : : : : : : : : : : : :
Db 379 KTCNSSLT LKTHHVODSKMGFVINAIVSMAYGLHNQMQLCPGYAGLCDAMKPIDGRKLL 438

QY 308 INLQKVIDOCSSDGFSGTHKCHLNNSECMPTIKGLGFVLGAECTICAGFYHPGVLPVN 367
   : : | : | : | : | : | : | : | : | : | :
Db 439 ESLMKNTFTGVSGDTILFDNGSDSPRYETLNNFKEMG-----KDYFDYINVGSD 488

```

Qy	368	NFRRRGPDQHI-----SGSKTDVSEE-----AYCCLPCREG-----	398
Db	489	NGELKMDDDBWSKKNIIIRSVCEPCKGQIKVIRKGEVSCWTCCTCKENVEYVFDEYT	548
Qy	399	CPCP-----ADD-SCFCFQVEDKYLR-----LAIISFQGLCMLLDFVSMLVYHFHRAK	445
Db	349	CKACQLSGSWPTDLDLTCDCEIPVOYLURWGSDPEPIAAVFAFAGLLATLFVTVVFFIYRDTP	608
Qy	446	SIRASGLILLETILFGSLLLYFPVWTLYPEPSTFRICILLRWARLIGFATVGTVTLKLHR	505
Db	609	VKSSSELZYIILAGICIGLYLCTCLIAKPKQIYCYLQRIGIGLSPAMSFALVTNTNR	668
Qy	506	VLKV-----FLSRTAQRIPYMTGGRYMRMLAVILLVFWFLIGHTSSVCQN	557
Db	669	IARLAGSKKKICTPKPREMSACQ-----LVIAFIILI-----CIQ	704
Qy	552	LEKOISL-----ICQGYTSBHLIFNMCLIDRWDYMTAVA-EFULMGVYLCYAVRTV	603
Db	705	LGIIVAFIMEPPDIMHDYPISREYLYCNTNLGVVTFPLGNNGLLILISCTFYAFAKTRNV	764
Qy	604	PSAFHEPRYMAVAHVNELLIISAIFETIRFVLASRLQSOMLMLYFAHTHLTVTITGLLL	663
Db	765	PANFPEAKVIATMYTTCIIWLAFPIYF-----GSNYKIITMCFVSLSATVALGCMF	818
Qy	564	IPK-----FSHSSNNPRDDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLPEDIRD	716
Db	819	VPAVYIILAKPERNVRSAFTTSTVVRMHVGDCSSSAARSSSLVNLMKRRGSSGETLSS	878
Qy	717	ELKK-----LYAOLEYTKRKMITNNPHLOK--KRCSKKGLGRIMRBIT	759
Db	879	NGRSVTWAONEKSSRGOWHLWRSLTHINKKENPNQTAVIKPFPKSTESRGLAG-	932
Qy	760	EIPETVSRQCKEDKBGDHGTAKTALIRKNPPESSCNTCKSEETLKNRVFSLKKSHS	819
Db	933	-----AGAG-GSAGGVGATGGNCAGACGPGESPADGAPKALY-----	969
Qy	820	TYDHVRDOTESSSLPTESQEBETTENTSTLESGLKLTQKLKEDSEABESTESVPLVCKS	879
Db	970	-----DVAEEHFPAPARPSPSPISTLSHRAG-----SASRTDDDVPSLHSEPVARRS	1019
Qy	880	ASAHNLSSEKKKGCHPR-TSMLOKLSLVTSASAKECTLGLAG-----KTQTA	923
Db	1020	SSOGSLMEIOISSVVRTFTANTISELNMMLSLAAPSFGVGAFLCYSYLIPKEIQLPMTMTT	1079
Qy	924	GVBERTKSOKPLPKOKETNRHNSNDNETTKOPAPON--SNPAEPERPKPKSGIMLKQOR	981
Db	1080	FAE-----IQPLPAIEVTGGAOPAGAQAAGDAARESPAACEAAAANKPDLEEIALVTPP	1134
Qy	982	NPTANSDLNPQTKQKDNFDIGEYCPWEVYDLTFGPVPVSESK	1024
Db	1135	SP-----FROSDSGSTTNPSNPVSESALCJIPSSPK	1164

RESULT 10
US-08-660-148-5
; Sequence 5, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1212 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-660-148-5

Query Match 3.18; Score 197.5; DB 4; Length 1212;
Best Local Similarity 18.68; Pred. No. 1.8e-07;
Matches 207; Conservative 135; Mismatches 383; Indels 387; Gaps 48;

Qy 219 ETEWFGHLLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPPLYECE 273
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 DVKWFDDYLLKLPETNHRNP-----WFQEPWQHRFQCRLEGFPQENSKYNT-----CN 382

Qy 274 NG-----SYRPGWLVL-TLSAIVGL---QPNLVPFGRGV-----MKVDINLQ 311
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 SSSLTKTHVQDSKMGFVINAIIYSMAVGLHNMQSLCPGYAGLGDAMKPIDGRKLLESIM 442

Qy 312 KVIDIDQSSDGFSTGTHKCHLNNECEMPIKGLFVLGAYECICKAGFYVHGVLPVNFRR 371
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 KTNFTGVSGDILFDENGDSGRYEIMNFKEG-----KDYFDYINVGWDNGEL 492

Qy 372 RGPQHI-----SGSTKDVSEE-----AYVCLPCREG-----CPFC 402
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 493 KMDDDEVMSKKSIIIRSVCEPCEKQIKVIRKGEVSCWCWTCCKENEYVFDEVYCKAC 552

403 -----ADD-SPCFVQEDKYLR-----LAIISFGCLMLDLDFVSMVYVYHFRKAKSIRA 449
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
553 QLGSWPTDDLTCGDLIPVQYLRWGDPEIAAVVFACLGLLATLFTVVVFIILYRDTVPVKS 612

Qy 450 SGLILLETILFGSLLLYFPVVLVFEPSFTFCILLRWARLLGFATVYCTVTLKLRHLKV 509
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 613 SSRELCYIILAGICLGYLCTCLIAKPKQIYCYLQRIIGLSPAMSYLSALVTKTNRARI 672

Qy 510 -----FLSRTAQIRIPYMTGGRVWRMLAVILLVFWFLIGWTSVQCNLEKQ 555
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 673 LAGSKKKICTKKPREMSACAQ-----LVIAFILL-----CIQLGII 708

Qy 556 ISL-----IGOKTSDHLIFNCLIDRWDMYTAVA-EFIFLWGVYLCVAVRTVPSAF 607
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 709 VALFTMEPPDIHHDYPSIREVYLICNTNLGVWTPLGVNGLLILSCTFYAFKTRNVPANF 768

Qy 608 HEPRYMAVAVHNELIISAFITIFVLASRLQSDWMLMYFAHTLTVTVTIGILLIPK- 666
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 769 NEAKYIAFTMTTCLILWAFVPIYF-----GSNYKIITMCFVSLSATVALGCMFVKV 822

Qy 667 ---FSHSSNNPRDDIATE-AYEDELDMGRSGSYL--NSSINSANSESLDPEIDRDELKK 720
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 YIILAKPERNVSFTSTVYVMHVGDCSSSAASRSSSLVNLKRRCSSETLR----- 877

Qy 721 LYAOLEYKRRKKMTNNPHLQKRCCKGLGRSINRRITEIPETVSRQCSKEDKEGADHG 780
```

```
Db 878 -----YKDRRLAQHKSEI-----ECF----- 893
Qy 781 TAKGTALIRKNPPSSNGTSGKSEETLKNRVFSLKSHSTYDHRVDOTESESSSLPTESQE 840
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 894 TPKG-----SMNGGRA-----TWSSNGKSVTWAQN 920
Qy 841 EETTENTLESLSGKKLTQKL-----KEDSE-----AESTESVPLVCKSASAHNLS 886
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 EKSSR-----GOHLWQRLSIHINKKNPNQTAIVKPPPKSTESRGLGAGAGAGGSAG 972
Qy 887 SEKKTG-----HPRTSMLOKLSVIAASAEKTLGLAGKTOTAGVEERTKSQK 933
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 973 GVGATGGAGCAGGPGGPESPDAG--PKALYDVAEAEHFPAPA-----RPRSPS 1020
Qy 934 PLPKDKETNRNHSNDTETKDPAPQ--NSNPAEPRPKQKSGIMKO--QRYNPPTAN--SD 989
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1021 PI-----STLSHRAGSASRTDDVPSLHSEVPARSSSQGS--LMEQISSVYVTRTANISE 1074
Qy 990 LNPCTTQMKDNFDIGEVCPWEVYDLTFG---PVPSESKVQKHVSIVASEMEKNPTFSLKE 1046
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 LN-----SMMLSTAAP-----SPGVGAPLCSSYLIIPKEIQL-----PT----- 1107
Qy 1047 KSHHKPRAAEVCCQSNQKRIDKAEVCLWESQGSILEDEKLLISKTPVLPERAKEENG- 1105
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1108 -----TMTTFAETIQPLPAIEVTGGA 1127
Qy 1106 QPRAANVCAG-----QSEELPPKAVAKTENENLQIGHQEKTKSSSEENVRGSYNSSNNF 1161
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1128 QPAGAQAAGDAARESPAGPEAAAKPDLEELVAL-----TPSPFRDSVDGSGT- 1178
Qy 1162 QOPLTSRAEVCWPEFETPAQPNAGRSVALPAS 1193
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1179 -----TPNSPVSESALECPSS 1194

RESULT 11
US-08-486-270-10
Sequence 10, Application US/08486270
Patent No. 5807689
GENERAL INFORMATION:
APPLICANT: Daggett, Lorrie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,270
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
```

QY 930 KSQKPLPKDKETNRNHSNSDNTETKDPAPQ-NSNPAAEPRKPQKSGIMKQ-QRVNPTTA 986

Query Match 3.0%; Score 193.5; DB 3; Length 1212;
Best Local Similarity 18.5%; Pred. No. 4e-07;
Matches 207; Conservative 131; Mismatches 383; Indels 395; Gaps 47;

QY 219 ETEWFHGLRRKWRPHLHRRGNQPGRLGHSHWRKDGKSHFKWSPPYLECEGNSY- 277
Db 333 DVKWFDDYILKLRPETNHRNP-----WQEFWQHR-----FQCRLEAFPOENSKYN 378
QY 278 -----KPGWLW-TLSSAIYGL---OPNLVPERGV-----MKVD 307
Db 379 KTCNSSLTLTHHVDKMGFVINAIIYSLMAYGLHNMQSLCPGYAGLCDAMKPIDGRKLL 438
QY 308 INLOKVIDDCCSDGWFSGTHKHLNNSECMPKGLGFLVGLAYECICKAGFYHGPVLPVN 367
Db 439 ESLMKTNFTGSGDITLFDENGSDSPGRYETMNFKEG-----KDYFDYINVGSWD 488
QY 368 NFRRGPDQHI-----SSTKDVSEE-----AYVCLPCREG----- 398
Db 489 NGELKMDDEVMSKSNIRSVCEPCEKQIKVIRKEVSCCWTCTPCKENYVFDEYT 548
QY 399 CPFC-----ADD-SPCFVQEDKYL-----LAIISFOGLCMLLDFVSLVYVHFKAK 445
Db 549 CKACQLGSWPTDGLTGDLLPQVYLRWGDPEPIAAVVFACGLLATLFTVTVVFIYRDTP 608
446 SIRASGLILLETILFGSLLLYFPVVIYFEPSTFRCLLRMARLLGPATVYGTVTLKLR 505
609 WKSSSRELVCYIILAGICLGLCTFCIAKPKQIYCYLQIGLSPMSYSALVTKTNR 668
QY 506 VLKV-----FLSRTAORTPYMTGGVRMRLAVILLVWFVFLIGHTSSVCQN 551
Db 669 IARLAGSKKICTPKPRFNSACQ-----LVIAFILI-----CIQ 704
QY 552 LEKQISL-----IGQKTSDDLHFNKCLIDRMDYMTAVA-BEFLFLMGVILCYAVRTV 603
Db 705 LGIIVAFIMEPPDIMHDYPSIREVYICNTNLGVVTPPLGNGLLLSLCTFTAFKTRNV 764
QY 604 PSAPHEPRYMAVAVHNELIISAFHTIRFVILASRLQSDWMLMYFAHTLTVTVTIGLL 663
Db 765 PANPEAKYIAFTWYTCILWAPVIYF-----GSNYKIITWCFVSLSATVALOCMF 818
QY 664 IPK-----FSHSSNNPRDIATE-AYEDELMDGRSGYL--NSSINSAMSHSLDPDIRD 716
Db 819 VPKYIILAKPERNVRSAFTTSTVVRHVGDGKSSAARSSSSLVNLWKRSGSGETLR- 877
QY 717 ELKLYAQLKRYKKMTNNPHLQKRCCKGLGRSIMRRIETPETVSRQCKEDKEG 776
Db 878 -----YKDRUAQHKEI-----ECF----- 893
QY 777 ADHGTAKTALIRKNPESGNTGKSEETLKNRVFSLKSHSYDHRDQTERESSLPT 836
Db 894 -----TPKG-----SMNGGGA-----TWSSNGKSVT 916
837 ESQEEETENSTLESLSGKLTOKL-----KEDSE-----AESTESVPLVCKSASA 882
Db 917 WAQNEKSSR-----GOHLWORLSHIHNKENPNQTAIVKPPKSTESRGLGAGAGAG 968
QY 883 HNLSSKKTG-----HPRTSMLQKSLVITASAKETLGLAGKTQTAGVEERT 929
Db 969 GSAGGVGATGACGACGAGCGPSPDAG--PKALYDVAEEHFPAPA-----RP 1016
QY 930 KSQKPLPKDRETNRNHNSDNTETKDPAPQ-NSNPABEPRKPKQSGIMKO--QRVNPFTA 986
Db 1017 RSPSPI-----STLSHRAGSASRTDDVPSLHSEPVARSSSOGS-LMEQISSVVTFTA 1070
QY 987 N-SDLNPGTOMKDNFDIGVCPWEVYDLTPG----PVPSESKVKQHVSIIVASEMEKNTF 1042
Db 1071 NISELN-----SMMLSTAAP-----SPGVGAPICSSYLIPKEIQL-----PT- 1107
QY 1043 SLKEKSHHKPAAEVCCQSNQKIDKAECVCIWESOGSILEDEKLLSKTPVLPERAKEE 1102
Db 1108 -----TMTFFAIOPLPAIEV 1123
QY 1103 NGG-OPRAANVCAG-----QSEELPPKAVASKTENENLQICHOEKKTSSEENVRGSGNS 1157
Db 1124 TGGAQPAAGAAQGAADARESPAAGPEAAAKPDLEELVAL-----TPPSPRDSVDS 1175
QY 1158 SNNFQQLTSRAEVCVPWEFETPAQPNAGRSVALPAS 1193

Db 1176 GST-----TPNSPVSESALCIPSS 1194
RESULT 13
US-08-072-574-8
; Sequence 8, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN MATABOTROPIC GLUTAMATE RECEPTORS,
; NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,574
; FILING DATE: 19930604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-072-574-8
Query Match 3.0%; Score 193; DB 1; Length 1180;
Best Local Similarity 18.8%; Pred. No. 4.2e-07;
Matches 177; Conservative 116; Mismatches 402; Indels 248; Gaps 33;
QY 219 ETEWFHGLRRKWRPHLHRRGNQPGRLGHSHWRKDGKSHFKWSPPYLECEGNSY- 277
Db 333 DVKWFDDYILKLRPETNHRNP-----WQEFWQHR-----FQCRLEAFPOENSKYN 378
QY 278 -----KPGWLW-TLSSAIYGL---OPNLVPERGV-----MKVD 307
Db 379 KTCNSSLTLTHHVDKMGFVINAIIYSLMAYGLHNMQSLCPGYAGLCDAMKPIDGRKLL 438
QY 308 INLOKVIDDCCSDGWFSGTHKHLNNSECMPKGLGFLVGLAYECICKAGFYHGPVLPVN 367
Db 439 ESLMKTNFTGSGDITLFDENGSDSPGRYETMNFKEG-----KDYFDYINVGSWD 488
QY 368 NFRRGPDQHI-----SSTKDVSEE-----AYVCLPCREG----- 398
Db 489 NGELKMDDEVMSKSNIRSVCEPCEKQIKVIRKEVSCCWTCTPCKENYVFDEYT 548
QY 399 CPFC-----ADD-SPCFVQEDKYL-----LAIISFOGLCMLLDFVSLVYVHFKAK 445
Db 549 CKACQLGSWPTDGLTGDLLPQVYLRWGDPEPIAAVVFACGLLATLFTVTVVFIYRDTP 608
QY 446 SIRASGLILLETILFGSLLLYFPVVIYFEPSTFRCLLRMARLLGPATVYGTVTLKLR 505

```
Db 609 VYKSSRELVCYIILAGICIGYICTECLIAKPKQIYCYIQIRIGISPAWSYALVTKNR 668
QY 506 VLKV-----FLSRTAQRIPTYMTGGRVMRLAVILLVVFVFLIGWTSSVCQN 551
Db 669 IARILAGSKKICTPKPREMSACQ-----LVIAFLI-----CLQ 704
QY 552 LKQISL-----TQGGTSOHLIFNMCILDRDYMATAV-EFIFLLMGVYLCYAVRTV 603
Db 705 LGIIVALFMEPPDMDHIDYPSIREVLYLNTNLGVVTPPLGNGLLILSCTEYAFKTRNV 764
QY 604 PSAFHEPRYMAVAVHNELIISAIFHTIRVFLASRLQSDWMLLYFAHHTLTVTVTIGLLL 663
Db 765 PANFPEAKYIATMTTCKIWLAFVIYF-----GSNYKIITMCFSVSLSATVALGCMF 818
QY 664 IPK-----FSHSSNNPRDDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEDIRD 716
Db 819 VPKVYIILAKPERNVRSFTTSTVVRMHVGDCKSSAARSLLVNLKRRGSSGETLSS 878
QY 717 ELKK-----LYAOLEYKRRKMTNNPHLOK-----KRCSSKGLGRSIMRRIT 759
Db 879 NGKSVTWAQNKSSRGQHLWQLRSLTHINKENPNQTAVIKPPKSTESRGLGAG----- 932
QY 760 EIPETVSRQCKEDKADHGTAKGTALIRKNPPSSGNTGKSKEETLKNRVFSLKKSHS 819
Db 933 -----AGAG-GSAGGVGATGGAGACGPGPESPDAGPKALF----- 969
QY 820 TYDHRDOTESSSLPTSEQEBETTENSTLSGKLTOKLKEDESEASTESVPLVCKS 879
Db 970 -----DYAEAEHPAPARPPSPPIRTLSHRAG-----SASRTDDVPSLHSEPVARS 1019
QY 880 ASAHNLSSKKTGHRP-TSMLOKSLSVIASAKEKTLGLAG-----KTQTA 923
Db 1020 SSQGSIMEQISSVWTRFTANISELSMMLSTAAPSPGVGAPLCCSYLPKEIQLTPTTWT 1079
QY 924 GVEERTKSKOLPKPKETNRHNSNDTETKDPAPON--SNPAEPRKPKOKSGIMKQORV 981
Db 1080 FAE-----IQPLPALEVGTGAQAAGAAAGAESPAAGPEAAAKPDLEELVALTTP 1134
QY 982 NPTTANSLNPGTQMKDNFDIGEVCPWEVYDLTPGPVPSESK 1024
Db 1135 SP-----FRSDVDSGSTTPNSPVSESAICIPSSPK 1164
```

RESULT 14

US-08-687-289A-8
; Sequence 8, Application US/08687289A
; Patent No. 5981195

GENERAL INFORMATION:

APPLICANT: Fuller, Forrest H.
APPLICANT: Krapcho, Karen J.
APPLICANT: Hammerland, Lance G.
TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
TITLE OF INVENTION: IDENTIFYING COMPOUNDS ACTIVE AT
TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSEQ for Windows 2.0

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: July 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-687-289A-8

Query Match 3.0%; Score 190; DB 2; Length 1056;

Best Local Similarity 19.9%; Pred. No. 6.3e-07;
Matches 124; Conservative 97; Mismatches 241; Indels 162; Gaps 20;

```
QY 357 GRYPHCVPLVNNFRRGDPDQHSIGSTKDYSE-----AYVCLPCREG- 398
Db 498 GTWHEGVNLIDDKYKTQ---MKNKSGVVRVSCSEPCCLKQIKVIRKGEVSCCMCTACKENE 554
QY 399 -----CPFC-----ADSPCFVQEDKYL-----LAIISFOGCMLLDFVSMLVV 438
Db 555 YVDEFTKACDGLGWPNADLTGCEPIPVRYLEWNIPIIAFASCGIILVTLFVTLIF 614
QY 439 YHFRKAKSGLILLLETLFGSLLLYFPVVILYFEPSTFCILLRWARLGFATVYGT 498
Db 615 VLYRDTVPVKSSRELVCYIILAGIFLVGVCPPTLIATPTTSCYLRLLVGLSSAMCYSA 674
QY 499 VTLKLRHLVKV-----FLSRTAQRIPTYMTGGRVMRLAVILLVVFVFLIGW 544
Db 675 LVTNTRIARILAGSKKICTRKPRMSAWAQVIAIISVQLTLVTVTLIT----- 726
QY 545 TSSVQCNLEKOISLIGOGKTSDDLIFENCLIDRWDMYMTAVAEF---LFLMGVYLCYAV 600
Db 727 -----MEPPMPIL-----SYPSIKEVILCNTSNLGVVAPLGVNGLLIMCTTYAFKT 774
QY 601 RTVPSAFHEPRYMAVAVHNELIISAIFHTIRVFLASRLQSDWMLLYFAHHTLTVTVTIG 660
Db 775 RNVPAFNEAKYIAFTMTTCLWLAFVIYF-----GSNYKIITTCFAVSLSVTVVALG 828
QY 661 LLLLPK-----FSHSSNNPRDDI--ATEAYEDELDM-----GRSGSYL 696
Db 829 CMFTPKMYIIITAKPERNTIEEVRCSTAAAFKVAARATLRNSVSRKRSSSLGGSTGTP 888
QY 697 NSSINSAMSEHSLDPEDIRDELKLYLAQLEYKRRKMT---NNPHLOKRCSSK---GL 750
Db 889 SSISSSKSNSEDDPPQPERQKQOQPLALTQEQOQOQUTLPQOQSQOQPKQKQVIFGS 948
QY 751 GRSIMKRRIITEIPETVSRQCKEDKADHGTAKGTALIRKNPPSSGNTGKSKEETLKN 810
Db 949 G-----TVTFLSFLDFEPQKNMAHGNS-----THQNS 975
QY 811 VFSLKKSHTYDHRDQTESSSLPTSEQEBE---TTNNTSLGSLGKLLQKLEKDESEA 867
Db 976 LEAKSSDTL-----TRHQPLLPLOQGETDLDTVQETGLQGPVGGDQRPVEDPEEL 1028
QY 868 ESTESVPLVCKKSASAHNLSSSEKKT 891
Db 1029 ---SPALVSVSSSQSFVISGGGST 1048
```


Align seg 1/1 to: AB032962 from: 1 to: 4595

```
672 AsnAsnProArgAspAspIleAlaThrGluAlaThrGluAspGluLeuAs 688
|||||
2 AATAACCCAGAGATGATATGCTACAGAGCATATGAGGATGAGCTAGA 51
|||||
688 pMetGlyArgSerGlySerThrLeuAsnSerSerIleAsnSerAlaTrps 705
|||||
52 CATGGGCGGATCTGGATCTTACCTGAACAGCAGCTATCAATTACAGCCTGA 101
|||||
705 erGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLeu 721
|||||
102 GTGAGCACAGCTTGGATCCAGAGAGCATTCGGGACGAGCTGAAAAAAGCT 151
|||||
722 TyrAlaGlnLeuGluIleThrLysArgLysLysMetIleThrAsnAsnPr 738
|||||
152 TATGCCCACTGGANATATATAAAGAGAGAGATGATCAACAACACCC 201
|||||
738 oHisLeuGlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleM 755
|||||
202 CCACCTCCAGAAAAAGCGTGTCTCGAAGAGAGGCGCTAGGTCGTTCATCA 251
|||||
755 etArgArgIleThrGluIleProGluThrValSerArgGlnCysSerLys 771
|||||
252 TGAGACGCATTCGGAGATCCAGAGACAGTCCAGCCGGCAGTCTCTAA 301
|||||
772 GluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuIl 788
|||||
302 GAGGACAGAGGCGCGCCGACATGCGACAGCCAAAGGCACCTGCCCTCAT 351
|||||
788 eArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLysGluG 805
|||||
352 CAGGAAGAACCCCCAGAGTCTTCAGGGAACACAGGGAATCCCAAGGAG 401
|||||
805 luThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerThrTyr 821
|||||
402 AGACCTCGAAAAACCGAGTCTTCTCACTCAAGAAATCCACAGCACTTAT 451
|||||
822 AspHisValArgAspGlnThrGluGluSerSerSerLeuProThrGluSe 838
|||||
452 GACCACGTGAGAGACCAACAGGAGAGTCCAGTAGCCTACCCACAGAAAG 501
|||||
838 rGlnGluGluGluThrThrGluAsnSerThrLeuGluSerLeuSerGlyL 855
|||||
502 CCAAGAGAGGAGAGACACAGAAAAATCCACACTGGAAATCCCTGTCGGTA 551
|||||
855 yLysLeuThrGlnLysLeuLysGluAspSerSerGluAlaGluSerThrGlu 871
|||||
552 AAAAATAACAAAAAATCAAGAGAGAGAGGAGGCTGAGTCCAGGAG 601
|||||
872 SerValProLeuValCysLysSerAlaSerAlaHisAsnLeuSerSerGI 888
|||||
602 TCGGTCGCGTGTGTGCAAGTCCAGAGCGCTCAACACCTCAGCTCAGA 651
|||||
888 uLysLysThrGlyHisProArgThrSerMetLeuGlnLysSerLeuSerV 905
|||||
652 GAAGAAAACTGGGACCCACAGCAATCGATGTTACAGAAGTCTCTCAGTG 701
|||||
905 aIleAlaSerAlaLysGluLysThrLeuGlyLeuAlaGlyLysThrGln 921
|||||
702 TCATAGCAAGCGCCAGGAGAGAGACTCTTGGATTAGCTGGGAAAAACCCAA 751
|||||
922 ThrAlaGlyValGluGluArgThrLysSerGlnLysProLeuProLysAs 938
|||||
752 ACAGCAGAGTGTGGAAGACGCACTAATCCAGAAACCTTTGCCAAAGA 801
|||||
938 pLysGluThrAsnArgAsnHisSerAsnSerAspAsnThrGluThrLysA 955
|||||
802 TAAAGAGACAAAACAGAAATCACTCAAAATTCGTATAACACAGAGACTTAA 851
|||||
955 spProAlaProGlnAsnSerAsnProAlaGluProArgLysProGln 971
|||||
```

```
852 ATCCTGCCCCCCCCAAAACTCAAAATCCTGCGGAGGAGCCAAAGAAAGCCTCAG 901
|||||
972 LysSerGlyIleMetLysGlnGlnArgValAsnProThrThrAlaAsnSe 988
|||||
902 AATCTGGGATTATGAACACAAAGGGTCAACCCACCACTGCCAATTC 951
|||||
988 rAspLeuAsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1005
|||||
952 TGACCTGAACCCAGGACCAACCCAGATGAAGGACAACCTTTGACATTTGGG 1001
|||||
1005 luValCysProThrProGluValTyrAspLeuThrProGlyProValProSer 1021
|||||
1002 AGGTGTCTCTTGGGAGTTTATGACCTGACCCCTGCTGTGCTTCA 1051
|||||
1022 GluSerLysValGlnLysHisValSerIleValAlaSerGluMetGluLy 1038
|||||
1052 GAATCAAAAGTTCAAAGGACAGTATCTATTGTGGCTTCTGAAAATGGAGAA 1101
|||||
1038 sAsnProThrPheSerSerLeuLysGluLysSerHisLysLysProLysAla 1055
|||||
1102 AAACCCCACTTTTCTTAAAGGAGAAATCTCACCAAGCCTAAGGCAG 1151
|||||
1055 laGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVal 1071
|||||
1152 CTGAGCTTTGTCCAGCAATCCAATCAGAGCGCATAGATAAGGCTGAAGTA 1201
|||||
1072 CysLeuTrpGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuLe 1088
|||||
1202 TGCCTTTGGGAGAGCCAGCCAGTCCATTTTGGAAAGATGAGAAAGCTTTT 1251
|||||
1088 uIleSerLysThrProValLeuProGluArgAlaLysGluGluAsnGly 1105
|||||
1252 GATTTCCAAGACTCCAGTTCTCCAGAGAGGCAAAAGAGGAGAACGGAG 1301
|||||
1105 lYcGlnProArgAlaAlaAsnValCysAlaGlyClnSerGluGluLeuPro 1121
|||||
1302 GTCAGCTCTGTCAGCAATGTGTGTCTGGGAGAGCGAAGAACTGCC 1351
|||||
1122 ProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleGI 1138
|||||
1352 CCCAAAGCTGTAGCATCAAAACAGAGAATGAAATCTCAACCAATAGG 1401
|||||
1138 yHisGlnGluLysLysThrSerSerSerGluGluAsnValArgGlySerT 1155
|||||
1402 ACACCAAGAAAAAAGACATCTTCTCAGGAGAAATGTCGTGGCTCCT 1451
|||||
1155 yAsnSerSerAsnAsnPheGlnGlnProLeuThrSerArgAlaGluVal 1171
|||||
1452 ATAACTCAAGTAATAACTTCCAGCAACCTTTAACTACAGCAGCAGAGTT 1501
|||||
1172 CysProTrpGluPheGluThrProAlaGlnProAsnAlaGlyArgSerVa 1188
|||||
1502 TGTCTCTGGGAGTTTGAGACCCCAAGCTCAACCAATGCTGGGAAGAAGTGT 1551
|||||
1188 lAlaLeuProAlaSerSerAlaLeuSerAlaAsnLysIleAlaGlyProA 1205
|||||
1552 AGCTTTACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1601
|||||
1205 rgLysGluGluIleTrpAspSerPheLysVal 1215
|||||
1602 GGAAAGAGAGAGGCTGGGATAGTTTTTAAAGTG 1633
```

seq_name: gb_hgt20:AL355587

seq_documentation_block:

LOCUS AL355587 151218 bp DNA
DEFINITION Homo sapiens chromosome 10 clone RP11-561H23, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL355587
VERSION AL355587.8 GI:13660940
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

89130 TCATAATACATGTTTATAACATCTTCTAGGAAAGGTAGGATAGATC 89179
713 713
89180 TGTGGAGCAGAGGAGAGCCTGGATTATTTAAATAAATCTATGAGCACA 89229
713 713
89230 TATCAGCATCTGAAGCAAGCAGATGTTTTAGTAGTCAACGGACTATTAC 89279
713 713
89280 CCTTAAGCCTAATCTAATCATCATGTAACTTTATCAATGGTGCATGA 89329
713 713
89330 TTCTTCCATATACTACAGGATATGGTTCACCTATAACAATATGCTGT 89379
713 713
89380 GTTTTCTTCTGGTCCAGGCTCAGTCTGGTCCCTCACAACTTGAATGTACAA 89429
713 713
89430 AGTGCTTGTCTTGTGATGCAAGGTTAGAAATACATGAGTTTCGTCTTAAA 89479
713 713
89480 GTCCACACTTTATGTGAACAAACAAACAAAGAAATCAGAAAATACAAATC 89529
713 713
89530 AGAAAGTACAAATCAGAAGGTCTTGCTCCTTGAAGCAAGGTAGAATA 89579
713 713
89580 CGTGAGTTCATGCTTAAAGGTCCACACTTTATGTGAACAAACAAACAA 89629
713 713
89630 GAATCAGAAAGTACAAATGGTCAGAAAATCGACAAGTACATATGTCAGAG 89679
713 713
89680 AAAGCTGTAGACAGCTTGCCCAAAATTAGAGCCCAAGTTTACTGAATCC 89729
713 713
89730 AATACCATGTAGTGTCTTCTATGACTGGACACTAAATTTCTACACTTCT 89779
714 ile.ArgAspGluLeuLysLysLeuTyrAlaG1 724
89780 TTGAATTTGCTTTTGTCTGGCAGCAGCAGCTGCAAAAACCTCTATGCCCA 89829
724 LeuGluLeuTyrLysArgLysLysMetileThrAsnAsnProHisLeuG 741
89830 ACTGGAATATATAAAGAAAGAAAGATGATCACAAACACACCCACCTCC 89879
741 InLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgArg 757
89880 AGAAAAGCGGTGCTCGAAGAGGCGCTAGGTCTTCCATCATGAGAGCG 89929
758 IleThrGluIleProGluThrValSerArgGlnCysSerLysGluAspLys 774
89930 ATTACGAGATCCCAGACAGACTACCGCGCAGTGTCTTAAGAGAGCAAA 89979
774 sGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuIleArgLysA 791
89980 CGAGGCGCGCAGCAGTGGCAGAGCCAAAGGCACCTGCCCTCATCAGGAAGA 90029
791 snProProGluSerSerGlyAsnThrGlyLysSerLysGluGluThrLeu 807
|||||

90030 ACCCCCCAGAGTCTTTCAGGGAACACAGGAAATCCAAAGGAGGAGCCCTG 90079
808 LysAsnArgValPheSerLeuLysLysSerHisSerThrTyrAspHisVa 824
|||||
90080 AAAAACCGAGTCTTCTCACTCAAGAAATCCACAGCAGCTTATGACCAGCT 90129
824 lArgAspGlnThrGluGluSerSerSerLeuProThrGluSerGlnGluG 841
|||||
90130 GAGAGACCAACGGAAGAGTCCAGTAGCCTACCCACAGAAAGCCAGAGG 90179
841 luGluThrThrGluAsnSerThrLeuGluSerLeuSerGlyLysLysLeu 857
|||||
90180 AGGAGACAAACAGAAATTCACACTGGAATCCCTGTGCGGTAAAAACTA 90229
858 ThrGlnLysLeuLysGluAspSerGluAlaGluSerThrGluSerValPr 874
|||||
90230 ACACAAAACCTAAAGAGACAGCAGGCTGAGTCCACGGAGTCGGTGCC 90279
874 leuValCysLysSerAlaSerAlaHisAsnLeuSerSerGluLysLysT 891
|||||
90280 GTTGTGTGCAAGTCAGCAAGCGCTCACAACTCAGCTCAGAGAAAGAAA 90329
891 hrGlyHisProArgThrSerMetLeuGlnLysSerLeuSerValIleAla 907
|||||
90330 CTGGCAGCCACGAACTGATGTTACAGAGTCTCTCAGTGTCTATAGCA 90379
908 SerAlaLysGluLysThrLeuGlyLeuAlaGlyLysThrGlnThrAlaG1 924
|||||
90380 ACGGCAAGGAGAAAGCTCTTGATTTAGCTGGGAAACCCAAACAGCAGG 90429
924 yValGluGluArgThrLysSerGlnLysProLeuProLysAspLysGluT 941
|||||
90430 TGTGAAGAAAGCACTAAATCCCAAGACCTTTGCCAAAGATAAAGAGA 90479
941 hrAsnArgAsnHisSerAsnSerAspAsnThrGluThrLysAspProAla 957
|||||
90480 CAACAGAAATCACTCAATTTCTGATAACACAGAGACTAAAGATCTCTGCC 90529
958 ProGlnAsnSerAsnProAlaGluProArgLysProGlnLysSerG1 974
|||||
90530 CCCCAAAACCTCAATCTCCGGAGGAGCCAGAAAGCCTCAGAAAATCTGG 90579
974 yIleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspLeuA 991
|||||
90580 GATTATGAACAACAAGGGTCAACCCACCACTGCCAATTTCTGACCTCA 90629
991 snProGlyThrThrGlnMetLysAspAsnPheAspIleGlyGluValCys 1007
|||||
90630 ACCCAGGCAACCCAGATGAAGGACAACTTTGACATTTGGGAGGTGTGT 90679
1008 ProTTPGluValTyrAspLeuThrProGlyProValProSerGluSerIy 1024
|||||
90680 CCTTGGGAGGTTTATGACCTGACCCCTGTCTGTGCTTTCAGAAATCAA 90729
1024 sValGlnLysHisValSerIleValAlaSerGluMetGluLysAsnProT 1041
|||||
90730 AGTTCAAAGCACGTATCTATTGTGGCTTTCGAAATGGAGAAAAACCCCA 90779
1041 hrPheSerLeuLysGluLysSerHisLysProLysAlaAlaGluVal 1057
|||||
90780 CTTTTCCTTAAAGGAGAAATCTCACCAAGCCTTAAGGCAGCTGAGGTT 90829
1058 CysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluValCysLeuTr 1074
|||||
90830 TGTACGCAATCCAATCAGACGCGCATAGATGAAGCTGAAGTATGCTTTG 90879
1074 pGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuLeuIleSerL 1091
|||||
90880 GGAGAGCCAGGCCAGTCCATTTTGGAGAGATGAGAAAGCTTTTGTATTCCA 90929
1091 ysThrProValLeuProGluArgAlaLysGluGluAsnGlyGlyGlnPro 1107
|||||
90930 AGACTCCAGTTCTCCAGAGAGGCGCAAGAGAGAGACGAGGAGTACGCT 90979

1108 ArgAlaAlaAsnValCysAlaGlyClnSerGluGluLeuProProLysAl 1124
|||||
90980 CGTGCAGCAATGTGTGTGGCAGACGAGCAAGTGCCTCCCAAGC 91029
|||||
1124 aValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleGlyHisGlnG 1141
|||||
91030 TGTAGCATCAAAACAGACAGATGAATCTCAACCAATAGACACACAGG 91079
|||||
1141 LuLysLysThrSerSerSerGluGluAsnValArgGlySerTyrAsnSer 1157
|||||
91080 AAAAAACACATCTTCTCTCAGGAGAAATGTGCGTGGCTCTATAACTCA 91129
|||||
1158 SerAsnAsnPheGlnGlnProLeuThrSerArgAlaGluValCysProTr 1174
|||||
91130 AGTAATAACTTCCACGACACCTTTAATCATCACGACGAGAGGTTTGTCTTG 91179
|||||
1174 pGluPheGluThrProAlaGlnProAsnAlaGlyArgSerValAlaLeuP 1191
|||||
91180 GGAGTTTGAGACCCAGCCAGCTCAACCAATGCTGGAAGAAGTGTACTTTAC 91229
|||||
1191 roAlaSerSerAlaLeuSerAlaAsnLysIleAlaGlyProArgLysGlu 1207
|||||
91230 CTGCCTCTTCTGCTTAAGTGCATAAGATAGACGAGGCCCTAGGAAGAA 91279
|||||
1208 GluLeuTrpAspSerPheLysVal 1215
|||||
91280 GAGATCTGGATAGTTTAAAGTG 91303
|||||

seq_name: gb_htg20:AL355542

seq_documentation_block:

LOCUS AL355542 185146 bp DNA HTG 13-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-257J14, *** SEQUENCING IN
PROGRESS ***, 19 unordered pieces.

ACCESSION AL355542

VERSION AL355542.9 GI:13624986

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 185146)

Plumb,B.

Direct Submission

Submitted (12-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Apr 14, 2001 this sequence version replaced gi:13624506.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA257J14

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 175310 bases at least Q40

Consensus quality: 178563 bases at least Q30

Consensus quality: 180883 bases at least Q20

Insert size: 183346; sum-of-contigs

Insert size: 145596; 18.0% error; agarose-fp

Quality coverage: 4.36x in Q20 bases; sum-of-contigs Quality

coverage: 7.64x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 19 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* runs of N, but the exact sizes of the gaps are unknown.

* runs of N, but the exact sizes of the gaps are unknown.

* runs of N, but the exact sizes of the gaps are unknown.

* runs of N, but the exact sizes of the gaps are unknown.

* runs of N, but the exact sizes of the gaps are unknown.

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 11771: contig of 11771 bp in length
11772 11871: gap of 100 bp
11872 61631: contig of 49760 bp in length
61632 61731: gap of 100 bp
61732 79327: contig of 17596 bp in length
79328 79427: gap of 100 bp
79428 83309: contig of 3882 bp in length
83310 83409: gap of 100 bp
83410 109528: contig of 26119 bp in length
109529 109628: gap of 100 bp
109629 112808: contig of 3180 bp in length
112809 112908: gap of 100 bp
112909 118905: contig of 5997 bp in length
118906 119005: gap of 100 bp
119006 133743: contig of 14738 bp in length
133744 133843: gap of 100 bp
133844 143213: contig of 9370 bp in length
143214 143313: gap of 100 bp
143314 158185: contig of 14872 bp in length
158186 158285: gap of 100 bp
158286 162061: contig of 3776 bp in length
162062 162161: gap of 100 bp
162162 164747: contig of 2586 bp in length
164748 164847: gap of 100 bp
164848 167658: contig of 2811 bp in length
167659 167758: gap of 100 bp
167759 170009: contig of 2251 bp in length
170010 170109: gap of 100 bp
170110 173787: contig of 3678 bp in length
173788 173887: gap of 100 bp
173888 176350: contig of 2463 bp in length
176351 176450: gap of 100 bp
176451 179441: contig of 2991 bp in length
179442 179541: gap of 100 bp
179542 182272: contig of 2731 bp in length
182273 182372: gap of 100 bp
182373 185146: contig of 2774 bp in length.

FEATURES

source

1. 185146
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-257J14"
/clone_lib="RPCI-11.1"

misc_feature

1. 11771
/note="assembly_fragment:00155
fragment_chain:1"

misc_feature

11872..61631
/note="assembly_fragment:00988
fragment_chain:1"

misc_feature

61732..79327
/note="assembly_fragment:01331
fragment_chain:1"

misc_feature

79428..83309
/note="assembly_fragment:03102
fragment_chain:1"

misc_feature

83410..109528
/note="assembly_fragment:02919
fragment_chain:1"

misc_feature

109629..112808
/note="assembly_fragment:01757
fragment_chain:1"

misc_feature

112909..118905
/note="assembly_fragment:01162
fragment_chain:1"

misc_feature

119006..133743
/note="assembly_fragment:01246
fragment_chain:1"

misc_feature

133844..143213
/note="assembly_fragment:01266
fragment_chain:1"

```
misc_feature      fragment_chain:1"
143314. .158185
/note="assembly_fragment:03087
fragment_chain:1"
158286. .162061
/note="assembly_fragment:00095"
162162. .164747
/note="assembly_fragment:00400"
164848. .167658
/note="assembly_fragment:00509"
167759. .170009
/note="assembly_fragment:01477"
170110. .173787
/note="assembly_fragment:01903"
173888. .176350
/note="assembly_fragment:02102"
176451. .179441
/note="assembly_fragment:02396"
179542. .182272
/note="assembly_fragment:02514"
182373. .185146
/note="assembly_fragment:03322"
BASE COUNT 50876 a 38769 c 39206 g 54480 t 1815 others
ORIGIN

alignment_scores:
  Quality: 2435.00      Length: 502
  Ratio: 4.909          Caps: 1
  Percent Similarity: 98.805  Percent Identity: 97.410

alignment_block:
US-09-775-181-2 x AL355542/rev ..
Align seg 1/1 to reverse of: AL355542 from: 1 to: 185146

715 ArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluIleTyrLysArgLy 731
:::
164477 CAGGACGAGTGAAGAACTCTATGCCCACTTGAATAATATATAAGAAA 164428
731 sLysMetIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysL 748
164427 GAAGATGATCACAACACACCCCACTCCAGAAAAGCGGTGCTCGAAGA 164378
748 yGlyLeuGlyArgSerIleMetArgArgIleThrGluIleProGluThr 764
164377 AGGCCTAGTCTGCTCATCATGAGACGATTACGCGAGATCCCGAGAGACA 164328
765 valSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisGlyTh 781
164327 GTACGCCGCGAGTGTCTAAAGAGGACAAGAGGCGCCGCCACCATGGCAC 164278
781 rAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyA 798
164277 AGCCAAAGGCACTGCCCTCATCAGGAAGAACCCCCCAGAGTCTTCAGGA 164228
798 snThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeu 814
164227 ACACAGGGAATCCAAAGGAGGAGACCCCTGAAAAACCGAGTCTCTCACTC 164178
815 LysLysSerHisSerThrTyrAspHisValArgAspGlnThrGluGluSe 831
164177 AGAAAAATCCACAGCAGCACTTATGACCACTGAGAGACCAACGGAAGATC 164128
831 rSerSerLeuProThrGluSerGlnGluGluThrGluThrGluAsnSerT 848
164127 CAGTACCTTACCACAGAAGCCAGAGGAGGAGACACAGAAAAATTCCTCA 164078
848 hrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAsp 864
164077 CACTGGAATCCCTGTGCGGTAAAAAACTAACACAAAAAACTAAAAAGAGAC 164028
865 SerGluAlaGluSerThrGluSerValProLeuValCysLysSerAlase 881
```

```
164027 AGCGAGGCTGAGTCCACGGAGTCGGTCCGTTGGTGTGCAAGTCAGCAAG 163978
881 rAlaHisAsnLeuSerSerCyluLysLysThrGlyHisProArgThrSerM 898
163977 CGCTCACACCTCAGCTCAGAGAAGAAACTGGGCACCCACGAGCAATCGA 163928
898 etLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeu 914
163927 TGTACAGAAGTCTCTCAGTGTCTATAGCAAGCCCAAGGAGAGACTCTT 163878
915 GlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSe 931
163877 GGATTAGCTGGGAAAAACCCACACAGCAGGTGTGGAAGACGCACATAATC 163828
931 rGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnS 948
163827 CCAGAAACCTTTGCCAAAAGATAAAGAGACAAACAGAAATCACTCAAAAT 163778
948 erAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAla 964
163777 CTGATAACACAGAGACTAAGATCCTGCCCCCAAACTCAAATCCTGCG 163728
965 GluGluProArgLysProGlnLysSerGlyIleMetLysGlnGlnArgVa 981
163727 GAGGAGCCAAAGAGCCTCAGAAATCTGGGATTATGAAACAACAAAGGTC 163679
981 lAsnProThrThrAlaAsnSerAspLeuAsnProGlyThrThrGlnMetL 998
163678 .AACCCACCACTGCCAATTTCTGACCTGAACCCAGCACCACCCAGATGA 163629
998 ysAspAsnPheAspIleGlyGluValCysProTyrGluValTyrAspLeu 1014
163628 AGGACAACCTTTGACATTGGGAGGTGTCTCTTGGAGGTTTATGACCTG 163579
1015 ThrProGlyProValProSerGluSerLysValGlnLysHisValSerI 1031
163578 ACCCTGTGCTGTGCTTTCAGAAATCAAAAGTTCAAAAGCAGGTATCTAT 163529
1031 eValAlaSerGluMetGluLysAsnProThrPheSerLeuLysGluLysS 1048
163528 TGTGGCTTCTGAATGGAGAAAACCCCACTTTTCTTAAAGGAGAAAT 163479
1048 erHisLysProLysAlaAlaGluValCysGlnGlnSerAsnGlnLys 1064
163478 CTCACCACAAGCCTAAGGCAGGTGATGTTGTTCAGCAATCAATCAATAG 163429
1065 ArgIleAspLysAlaGluValCysLeuTyrGluSerGln. GlyGlnSerI 1081
163428 CGCATAGATAAGGCTGAAGTATGCCTTTGGGAGAGCCAAAGGCCAGTCCA 163379
1081 leLeuGluAspGluLysLeuLeuIleSerLysThrProValLeuProGlu 1097
163378 TTTGGGAAGATGAGAAGCTTTTGATTTCCAGACTCCAGTTCTCCAGAG 163329
1098 ArgAlaLysGluGluAsnGlyGlyGlnProArgAlaAlaAsnValCysAl 1114
163328 .AGGCCAAAAGAGAGAACGAGAGTCAGCTCTGCGAGCAATGTGTGTCG 163279
1114 adlyGlnSerGluGluLeuProProLysAlaValAlaSerLysThrGluA 1131
163278 TGGCAG .AGCGAAGAACTGCCCCCAAGCTGTAGCATCAAAAACAGAGA 163230
1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSerSer 1147
163229 ATGAAAAATCAACCAATAGGACACCCAGAAAAAAGACATCTTCTCT 163180
1148 GluGluAsnValArgGlySerTyrAsnSerSerAsnAsnPhelGlnPr 1164
163179 GAGGAGAATGTCCGTGGCTCTCTATAACTCAAGTAATACTTCCAGCAACC 163130
1164 oLeuThrSerArgAlaGluValCysProTyrGluPheGluThrProAlaG 1181
```



```
653 AAGAGGAAATCTCACCACAGCCTAAGCAGCTGAAGTTTGTGACCAAT 702
1061 exAsnGlnLysArgIleAspLysAlaGluValCysLeuTrpGluSerGln 1077
      |||||
703 CCAATCAGAAAGTCACAGATAAGGCTGAAGTAGCTTGGGGGATCCAA 752
1078 GlyGlnSerIleLeuGluaspGluLysLeuLeuIleSerLysThrProVa 1094
      |||||
753 GGCCAGTCCATTGGAAGATGAGAACATTTTCATTTCGAAGACTCCAGT 802
1094 LLeuProGluArgAlaLysGluGlnGlyGlnProArgAlaAlaA 1111
      |||||
803 TCTCCAGAGAGGGCAAGAGAGGAGGAGGAGGCTCAGCTCAGGAGCCA 852
1111 snValCysAlaGlyGlnSerGluGluLeuProProLysAlaValAlaSer 1127
      |||||
853 AAGTGTGTCTGGGCAGAGTGAAGAACTGCCCCCAAGCTGTAGCATCA 902
1128 LysThrGluAsnGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysTh 1144
      |||||
903 AAACACAGAGAATGAAATCTCAACCAAAATAGGACACAGCAAAAAAGAC 952
1144 rSerSerSerGluGluAsnValArgGlySerTyrAsnSerSerAsnAsn 1161
      |||||
953 A...TCTTTGAGGAGAAATGTGGTGCTCTCTATAACTCAAGTAATACT 999
1161 heGlnGlnProLeuThrSerArgAlaGluValCysProTrpGluPheGlu 1177
      |||||
1000 TCCAGCAACCTTTAATGTCCAGCAGCAGAGGTGTGCCCTTGGGAGTTTGA 1049
1178 ThrProAlaGlnProAsnAlaGlyArgSerValAlaLeuProAlaSerSe 1194
1050 ACCCCAGCTCAACCAATCTCGAAGAACTGAGCTTTACCTGTTTCTTC 1099
1194 rAlaLeuSerAlaAsnLysIleAlaGlyProArgLysGluGluIleTrpA 1211
1100 TGCTCTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGG 1149
1211 spSerPheLysVal 1215
1150 ATACTTTTAAAGTG 1163
```

seq_name: gb_hgt21:ALJ358932

```
seq_documentation_block:
LOCUS      ALJ358932      111117 bp      DNA      HTG      23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN
PROGRESS ***
VERSION    ALJ358932.3      GI:9930957
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 111117)
AUTHORS    Direct Submission
TITLE      Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk clone
requests: clonerequest@sanger.ac.uk
COMMENT    On Aug 27, 2000 this sequence version replaced gi:9926667.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: BA395P8
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
```

Consensus quality: 90261 bases at least Q40
Consensus quality: 99226 bases at least Q30
Consensus quality: 104490 bases at least Q20
Insert size: 108717; sum-of-contigs
Insert size: 178772; 24.0% error; agarose-fp
Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality
coverage: 1.42x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2287: contig of 2287 bp in length
* 2288 2387: gap of 100 bp
* 2388 6317: contig of 3930 bp in length
* 6318 6417: gap of 100 bp
* 6418 9725: contig of 3308 bp in length
* 9726 9825: gap of 100 bp
* 9826 16607: contig of 6782 bp in length
* 16608 16707: gap of 100 bp
* 16708 20159: contig of 3452 bp in length
* 20160 20259: gap of 100 bp
* 20260 23641: contig of 3382 bp in length
* 23642 23741: gap of 100 bp
* 23742 27393: contig of 3658 bp in length
* 27400 27499: gap of 100 bp
* 27500 32754: contig of 5255 bp in length
* 32755 32854: gap of 100 bp
* 32855 35240: contig of 2386 bp in length
* 35241 35340: gap of 100 bp
* 35341 40948: contig of 5608 bp in length
* 40949 41048: gap of 100 bp
* 41049 43758: contig of 2710 bp in length
* 43759 43858: gap of 100 bp
* 43859 48684: contig of 4826 bp in length
* 48685 48784: gap of 100 bp
* 48785 52405: contig of 3621 bp in length
* 52406 52505: gap of 100 bp
* 52506 56282: contig of 3777 bp in length
* 56283 56382: gap of 100 bp
* 56383 59669: contig of 3287 bp in length
* 59670 59769: gap of 100 bp
* 59770 63200: contig of 3431 bp in length
* 63201 63300: gap of 100 bp
* 63301 68131: contig of 4831 bp in length
* 68132 68231: gap of 100 bp
* 68232 70678: contig of 2447 bp in length
* 70679 70778: gap of 100 bp
* 70779 72875: contig of 2097 bp in length
* 72876 72975: gap of 100 bp
* 72976 79043: contig of 6068 bp in length
* 79044 79143: gap of 100 bp
* 79144 88143: contig of 9000 bp in length
* 88144 88243: gap of 100 bp
* 88244 91654: contig of 3411 bp in length
* 91655 91754: gap of 100 bp
* 91755 102397: contig of 10643 bp in length
* 102398 102497: gap of 100 bp
* 102498 106184: contig of 3687 bp in length
* 106185 106284: gap of 100 bp
* 106285 111117: contig of 4833 bp in length.

FEATURES
source
1. .111117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-395P8"
/clone_lib="RPC1-11.2"
1. .2287
misc_feature


```
/note="assembly_fragment:00042
fragment_chain:1"
misc_feature 2388..6317
/note="assembly_fragment:00050
fragment_chain:1"
misc_feature 6418..9725
/note="assembly_fragment:00150
fragment_chain:2"
misc_feature 9826..16607
/note="assembly_fragment:00506
fragment_chain:2"
misc_feature 16708..20159
/note="assembly_fragment:00247
fragment_chain:3"
misc_feature 20260..23641
/note="assembly_fragment:00502
fragment_chain:3"
misc_feature 23742..27399
/note="assembly_fragment:00625
fragment_chain:4"
misc_feature 27500..32754
/note="assembly_fragment:00510
fragment_chain:4"
misc_feature 32855..35240
/note="assembly_fragment:00035"
35341..40948
/note="assembly_fragment:00100"
41049..43758
/note="assembly_fragment:00117"
43859..48684
/note="assembly_fragment:00200"
48785..52405
/note="assembly_fragment:00227"
52506..56282
/note="assembly_fragment:00235"
56383..59669
/note="assembly_fragment:00253"
59770..63200
/note="assembly_fragment:00303"
63301..68131
/note="assembly_fragment:00304"
68232..70678
/note="assembly_fragment:00421"
70779..72875
/note="assembly_fragment:00444"
72976..79043
/note="assembly_fragment:00512"
79144..88143
/note="assembly_fragment:00518"
88244..91634
/note="assembly_fragment:00562"
91755..102397
/note="assembly_fragment:00573"
102498..106184
/note="assembly_fragment:00584"
106285..111117
/note="assembly_fragment:00600"
BASE COUNT 32506 a 20474 c 20158 g 35546 t 2433 others
ORIGIN

alignment_scores:
  Quality: 1607.00      Length: 301
  Ratio: 5.339          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-181-2 x AL358932 ..
|||||
1 MetGlyAlaMetAlaTyrProLeuLeuLeuLeuLeuAlaGlnLe 17
|||||

Align seg 1/1 to: AL358932 from: 1 to: 111117
```

```
83403 ATGGAGCCCATGGCTTACCCCTTACTCCTCTGCTCTCTGCTGCTCAGCT 83452
17 uGlyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAsps 34
|||||
83453 GGGATTGGAGCTGTTGGCGCCAGCGCGACCCCAAGGAGCGCGGATT 83502
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
|||||
83503 CCCCTCGAGAGAGAGACCCGGAAGGGGAAGCGCACGCCAGCAGCGCGGT 83552
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG1 67
|||||
83553 CGACCTCTGCCTCGGACTCCTCGGCTCCCTGGAGCCGCTCCACCGATGG 83602
67 yThrIleLeuAlaGlnLysLeuAlaGluValProMetAspValAlas 84
|||||
83603 CACCATCTTGGCGCAGAAACTCGCCGAGGAGTGCCTCATGGACGTGGCCT 83652
84 erTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCySer 100
|||||
83653 CTTACCTCTACACCGGGGACTCCCAACGAGCTGAAGCGAGCCCACTGCTC 83702
101 GlyArgTyrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSe 117
|||||
83703 GCGCGCTACGAGTTGGCGGGCTCCCGGGGAAGTGGCCAGCCCTGGCCAG 83752
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
|||||
83753 CGCGCACCCCTCCTTGCACCGGGGCTGGACACACTGACACACGCCACCA 83802
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
|||||
83803 ACTTCCTCAACGTGATGTCAGAGCAATAAGTCGCGGAGCAACTTG 83852
151 GlnAspAspLeuAspTyrTrpGlnAlaLeuValTrpSerLeuGluG1 167
|||||
83853 CAGGACGACTGGATTGGTACCAGCGCTGGTGTGGAGCCCTCTGGAGGG 83902
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeus 184
|||||
83903 CGAGCCGAGCATCTCCGGGGGGCCATCACCCTTCAGCACCGATTGCTGT 83952
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluSerArg 200
|||||
83953 CCGCACCGGGCCCAAGGCTCTTCTCCAGGCCACGCGGAGGAGCGCGC 84002
201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
|||||
84003 ATCTGTCTCCAAGACCTGTCTCTCCGCAACCCACCTGGCCCAACGCCAC 84052
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
|||||
84053 TCTGGAGACGAGTGGTTCACAGGCTCCGCGCAAGTGGAGGCCCACT 84102
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
|||||
84103 TACACCGCGCGGCCCAATAACAGGGGCCCGGGGCTGGGCCACAGCTGG 84152
251 ArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerPr 267
|||||
84153 CGGCGCAAGGACGGGCTCGCGGGGACAAGAGCCACTTCAAGTGGTCTCC 84202
267 oProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
|||||
84203 GCCTTATCTGGAGTGGAGAACGGGAGTTACAAGCCCGGGTGGCTGGTTA 84252
284 hrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluPhe 300
|||||
84253 CTCCTTCTCTGCCATCTACGGGTTGACGCTTAACCTGGTCCCGGAATTC 84302
301 Arg 301
|||||
84303 AGG 84305
```

```
seq_name: gb_htg20.AL139821
seq_documentation_block:
LOCUS AL139821 199526 bp DNA HTG 17-MAR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-80K21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL139821
VERSION AL139821.7 GI:13396382
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Plumb.B.
Direct Submission
Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:10185474.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA80K21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 197300 bases at least Q40
Consensus quality: 198299 bases at least Q30
Consensus quality: 198676 bases at least Q20
Insert size: 199026; sum-of-contigs
Insert size: 14047; 11.6% error; agarose-fp
Quality coverage: 7.42x in Q20 bases; sum-of-contigs Quality
coverage: 10.32x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 37012: contig of 37012 bp in length
* 37013 37112: gap of 100 bp
* 37113 163560: contig of 126448 bp in length
* 163561 163660: gap of 100 bp
* 163661 171563: contig of 7903 bp in length
* 171564 171663: gap of 100 bp
* 171664 186786: contig of 15123 bp in length
* 186787 186886: gap of 100 bp
* 186887 196057: contig of 9171 bp in length.
* 196058 196157: gap of 100 bp
* 196158 199526: contig of 3369 bp in length.
*
* Location/Qualifiers
*
* 1. .199526
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="10"
* /clone="RP11-80K21"
* /clone_lib="RPC1-11.1"
* /clone="37012"
* /note="assembly_fragment:00668
fragment_chain:1
clone_end:T7
vector_side:left"
37113 - 163560
/note="assembly_fragment:02214
```

```
fragment_chain:1
clone_end:SP6
vector_side:right"
163661 - 171563
/note="assembly_fragment:00510"
171664 - 186786
/note="assembly_fragment:01538"
186887 - 196057
/note="assembly_fragment:01931"
196158 - 199526
/note="assembly_fragment:03255"
BASE COUNT 57428 a 38898 c 39349 g 63347 t 504 others
ORIGIN
alignment_scores:
Quality: 1607.00 Length: 301
Ratio: 5.339 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-775-181-2 x AL139821
Align seg 1/1 to: AL139821 from: 1 to: 199526
1 MetGlyAlaMetAlaTyrProLeuLeuLeuCysLeuLeuAlaGlnLe 17
97158 ATGGGAGCCATGCTTACCCCTTACTCCTCTGCTCTGCTGCTCAGCT 97207
17 uGlyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAspS 34
97208 GGGATTGGGAGCTGTGGCGCCAGCGCGCAGCCCAAGGACGCCGGATT 97257
34 exProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
97258 CCCCCTGAGAGAGAGAGCCCGAAGGGAGAGCCGACGCCAGCAGCGGGT 97307
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG1 67
97308 CGAGGCTCTGCTCGGACTCCTCGGCTCCTCGGAGCGCTCCACCGATGG 97357
67 yThrLeuLeuAlaGlnLysLeuAlaGluValProMetAspValAlaAs 84
97358 CACCATCTTGGCGCAGAAACTCGCGAGGAGGTGCCCATGGAGCTGGCCT 97407
84 exTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCysser 100
97408 CTTACTCTACACCGGGGACTCCACAGCTGAGCGAGCCCACTGCTCC 97457
101 GlyArgTyrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSe 117
97458 GCCCGCTACGAGTTGGCGGCTCGCGGGAAGTGGCCAGCCCTGGCCAG 97507
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
97508 CGCGCACCCCTCTCTTGCACCGGGCGCTGGACACACTGACACACGCCACA 97557
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
97558 ACTTCTCAACGTGTGTGTCAGAGCAATAAGTCGGGAGCAGAACTTG 97607
151 GlnAspAspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGlu1 167
97608 CAGGACGACCTGGATTGGTACCAGCGCTGGTGTGGAGCCCTTCTGGAGG 97657
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuS 184
97658 CGAGCCCGAGCATCTCCCGGGCGGCATCACCTTCAGACCCGATTCGCTGT 97707
184 exAlaProAlaProGlnValPheLeuGlnAlaThrArgGluSerArg 200
97708 CCGCACCGGGCCCCACAGGTCTTCTTCCAGGCCACCGCGAGGAGAGCCGC 97757
```

```
201 l1eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
97758 ATCTGCTCCAGACCTGCTCTCCGACCCACCTGGCCACGAC 97807
217 rLeuGlnThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
97808 TCTGGAGACCGAGTGTTCACGGCTCCGGCGGCAAGTGGAGGCCCACT 97857
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
97858 TACACCGCGCGGCCCAATCAGGGGCCCGGGCCCTGGGCCACAGCTGG 97907
251 ArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerPr 267
97908 CGGCGCAGGACGGCTCGCGGGGACAAAGACCACTTCAAGTGGTCTCC 97957
267 oProTrpLeuGluCysGluAsnGlySerTrpLysProGlyTrpLeuValT 284
97958 GCCTTACTGAGTGGCAGACGAGGAGTTCACAGCCCGGGTGGCTGGTTA 98007
284 hrLeuSerSerAlaIleTrpGlyLeuGlnProAsnLeuValProGluPhe 300
98008 CTCCTTCCTGCTGCATCTACGGGTTGACGCTAACCTGGTCCCGGAATTC 98057
301 Arg 301
|||
- 98058 AGG 98060
seq_name: gb_htg21:AL359974
seq_documentation_block:
LOCUS AL359974 143878 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-236A4, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION AL359974
VERSION AL359974.4 GI:9864577
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sims,S.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:9795088.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BB236A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: Plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 130224 bases at least Q40
Consensus quality: 135875 bases at least Q30
Consensus quality: 138890 bases at least Q20
Insert size: 141478; sum-of-contigs
Insert size: 205930; 23.5% error; agarose-fp
Quality coverage: 3.17x in Q20 bases; sum-of-contigs Quality
coverage: 2.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

```
* as soon as it is available and the accession number will
* be preserved.
1 12130: contig of 12130 bp in length
12131 12230: gap of 100 bp
12231 12274: contig of 8044 bp in length
20275 20374: gap of 100 bp
20375 23062: contig of 2688 bp in length
23063 23162: gap of 100 bp
23163 26227: contig of 3065 bp in length
26228 26327: gap of 100 bp
26328 33260: contig of 6933 bp in length
33261 33360: gap of 100 bp
33361 35571: contig of 2211 bp in length
35572 35671: gap of 100 bp
35672 38834: contig of 3163 bp in length
38835 38934: gap of 100 bp
38935 44371: contig of 5437 bp in length
44372 44471: gap of 100 bp
44472 47729: contig of 3258 bp in length
47730 47829: gap of 100 bp
47830 52657: contig of 4828 bp in length
52658 52757: gap of 100 bp
52758 54776: contig of 2019 bp in length
54777 54876: gap of 100 bp
54877 57327: contig of 2451 bp in length
57328 57427: gap of 100 bp
57428 60599: contig of 3172 bp in length
60600 60699: gap of 100 bp
60700 64861: contig of 4162 bp in length
64862 64961: gap of 100 bp
64962 71930: contig of 6969 bp in length
71931 72030: gap of 100 bp
72031 75429: contig of 3399 bp in length
75430 75529: gap of 100 bp
75530 78541: contig of 3012 bp in length
78542 78641: gap of 100 bp
78642 80961: contig of 2320 bp in length
80962 81061: gap of 100 bp
81062 83099: contig of 2038 bp in length
83100 83199: gap of 100 bp
83200 95484: contig of 12285 bp in length
95485 95584: gap of 100 bp
95585 97735: contig of 2151 bp in length
97736 97835: gap of 100 bp
97836 102251: contig of 4416 bp in length
102252 102351: gap of 100 bp
102352 131392: contig of 29041 bp in length
131393 131492: gap of 100 bp
131493 137581: contig of 6089 bp in length
137582 137681: gap of 100 bp
137682 143878: contig of 6197 bp in length.
FEATURES
Source
1..143878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RP13-13.1"
1..12130
/note="assembly fragment:01047
fragment_chain:1"
12231..20274
/note="assembly fragment:00007
fragment_chain:1"
20375..23062
/note="assembly fragment:01237
fragment_chain:1"
23163..26227
/note="assembly fragment:01103
fragment_chain:1"
26328..33260
/note="assembly fragment:00266
fragment_chain:1"
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
```

```
misc_feature 33361..35571
/note="assembly_fragment:01006
fragment_chain:1"
misc_feature 35672..38834
/note="assembly_fragment:00201
fragment_chain:2"
misc_feature 38935..44371
/note="assembly_fragment:01057
fragment_chain:2"
misc_feature 44472..47729
/note="assembly_fragment:00179
fragment_chain:2"
misc_feature 47830..52657
/note="assembly_fragment:00444
fragment_chain:3"
misc_feature 52758..54776
/note="assembly_fragment:00414
fragment_chain:3"
misc_feature 54877..57327
/note="assembly_fragment:00586
fragment_chain:4"
misc_feature 57428..60599
/note="assembly_fragment:00641
fragment_chain:4"
misc_feature 60700..64861
/note="assembly_fragment:00717
fragment_chain:5"
misc_feature 64962..71930
/note="assembly_fragment:01116
fragment_chain:5"
misc_feature 72031..75429
/note="assembly_fragment:00156"
fragment_chain:5"
misc_feature 75530..78541
/note="assembly_fragment:00233"
fragment_chain:5"
misc_feature 78642..80961
/note="assembly_fragment:00273"
fragment_chain:5"
misc_feature 81082..83099
/note="assembly_fragment:00279"
fragment_chain:5"
misc_feature 83200..95484
/note="assembly_fragment:00515"
fragment_chain:5"
misc_feature 95585..97735
/note="assembly_fragment:00654"
fragment_chain:5"
misc_feature 97836..102251
/note="assembly_fragment:00675"
fragment_chain:5"
misc_feature 102352..131392
/note="assembly_fragment:00719"
fragment_chain:5"
misc_feature 131493..137561
/note="assembly_fragment:00767"
fragment_chain:5"
misc_feature 137682..143878
/note="assembly_fragment:00852
clone_end:SP6
vector_side:right"
BASE COUNT 44271 a 25988 c 25788 g 45414 t 2417 others
ORIGIN

alignment_scores:
  Quality: 413.50      Length: 447
  Ratio: 1.854        Caps: 19
  Percent Similarity: 49.888      Percent Identity: 30.872

alignment_block:
US-09-775-181-2 x AL359974/rev ..
Align seg 1/1 to reverse of: AL359974 from: 1 to: 143878
372 ArqGlyProAspGlnHisThrSerGlySerThrLysAspValSerGluCl 388
89283 AGGGGTCGGGATCAGCATATTTTCAGAAAGTACAAAAGATGTGTCAAGA 89234
388 uAlaTrpValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
89233 AGCCATATGTCCTACCTTTCAGGAGGGGCTGCCCTTCTGTCTGATG 89184
```

```
405 spSerProCysPheValGlnGluAspLysTyrLeuArgLeuAlaIleIle 421
|||||
89183 ACAGCCCATGCTTCGTCAGGAAGATAAGTATTATTCAGACTTGCCATCATC 89134
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
89133 TCCTTCCAAGGCTGTGTATGCTGCTCGACTTCGTTAGCATGCTGTGGT 89084
438 lTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuLeuL 455
|||||
89083 CTACCACTTTCCAAAGCAAG..... 89062
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProValIle 471
|||
89061 .....GTAAACCCAGCAACCTG.....GTTATGATC 89035
472 LeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaAr 488
|||||
89034 CTGTATTAC..... 89026
488 gLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisA 505
|||||
89025 .....A 89025
505 rgValLeuLys ValPheLeuSerArgThrAlaGlnArgIleProTyrMe 521
|||
89024 GAGCAACCTCTTATCTTTTAAAGAAA.....CAGATTTCCTCAATCCA 88981
521 tThrGlyGlyArgValMetArgMetLeuAlaValIle..... 533
|||||
88980 AGTTGCTCCCTCTTAAGCAGCAGCTTTTCTACAATATGATTGGACTAGA 88931
534 ..LeuLeuValValPheTrpPhe..... 540
88930 ACATGAAGTTATATCTTTAGCTGACCAGAAAATAGAGATTAAACAT 88881
541 .....LeuIleGlyTrpTh 545
88880 GATACCTTTATTTTAACTTTACAGACTGTTCCCTGTTTGTAGCTGGAT 88831
545 rSerSerValCysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnG 562
|||||
88830 CTCCTAGTGCT.....T 88820
562 lYlYThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
|||
88819 CATTGACTTAAATGACTCTCCTTTACATCATGCTTTGGCCTCTAGAGAA 88770
579 AspTyrMetThrAlaValAlaGluPheLeuPheLeuLeuTrpGlyValTy 595
|||
88769 ACTTATCTACCAGCATCAATGATGATTATTAATTTTGTATCCAGAA 88720
595 rLeuCystYrAlaValArgThrValProSerAlaPheHisGluProArgT 612
|||||
88719 GCTTTGTTAT..... 88710
612 yrMetAlaValAlaValHisAsnGluLeuIleSerAlaIlePheHis 628
|||
88709 .....GTTTATTAT 88701
629 ThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrp...MetLe 644
|||
88700 APTTTTCTTTCTTCTAGTTGTGTTTAAATA.....TCGTAATGAC 88661
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
|||
88660 TCAACTATTTTTCAGGCATATCTTTCAGAGAAATGAATTTATTTATGACC 88611
661 euLeuLeu...IleProLysPheSerHisSerSerAsnProArgAsp 676
|||
88610 TTATATTGCTTACATAGTGTCCAGAAATATAACAATCCCGGGAAG 88561
```

```
677 AspIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGI 693
*****
88560 AACTTA.....CAGATTTTGGAGAGAACATGCTTTGGTTAGTCT 88517
|||
693 ySer_TyrLeuAsnSerIleAsnSerAlaTtp..... 704
|||
88516 AAGTAATCAACAGTAGTACATCCAGGAGCAGTGCTCATCCCTGTAATC 88467
|||
705 SerGluHls.....SerIle 709
|||
88466 TCAGCACATTGGGAGGCCAAGGCTGGAGAATTGCTTGAGGCCAGGAGTTT 88417
|||
709 u_AspProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeu 725
|||
88416 GAGACAGCCCTGACCAACATAGCAAGACCCCTGCTCTACAAAAAACA 88367
|||
726 GluIleTyrLysArgLysLysMetIleThrAsnAsn.ProHlsLeuGlnL 742
*****
88366 CAAATTTTAAAGCTAGCAATGCTGGTGATGCTATAGTCCAA 88317
|||
742 yLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgIle 758
*****
88316 CTATTTAGGAGCTGAGAAAGGAGTA.....TCTCTTGAGTCTAGAGTT 88273
|||
759 ThrGluIleProGluThrValSerArgGlnCys 769
|||
88272 CAAGACTACAGTACGCTATGATCATCACCACTGT 88240
|||
seq_name: gb_htg20:AL161654

seq_documentation_block:
LOCUS AL161654 159596 bp DNA HTG 11-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-59G22, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION AL161654
VERSION AL161654.8 GI:13620309
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton,J.
Direct Submission
Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerrequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:13567947.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA59G22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 157954 bases at least Q40
Consensus quality: 158428 bases at least Q30
Consensus quality: 158729 bases at least Q20
Insert size: 158896; sum-of-contigs
Quality coverage: 6.69x in Q20 bases; agarose-fp
coverage: 6.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

```
* as soon as it is available and the accession number will
* be preserved.
* 1 24276: contig of 24276 bp in length
* 24277 24376: gap of 100 bp
* 24377 31732: contig of 7356 bp in length
* 31733 31832: gap of 100 bp
* 31833 67342: contig of 35510 bp in length
* 67343 67442: gap of 100 bp
* 67443 70588: contig of 3146 bp in length
* 70589 70688: gap of 100 bp
* 70689 81295: contig of 10607 bp in length
* 81296 81395: gap of 100 bp
* 81396 111898: contig of 30503 bp in length
* 111899 111998: gap of 100 bp
* 111999 126181: contig of 14183 bp in length
* 126182 126281: gap of 100 bp
* 126282 159596: contig of 33315 bp in length.
FEATURES
Location/Qualifiers
1..159596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPC1-11.1"
1..24276
/note="assembly fragment:01962
fragment_chain:1
clone_end:T7
vector_side:left"
24377..31732
/note="assembly fragment:01615
fragment_chain:1"
31833..67342
/note="assembly fragment:02112
fragment_chain:1"
67443..70588
/note="assembly fragment:00781
fragment_chain:1"
70689..81295
/note="assembly fragment:00892
fragment_chain:1"
81396..111898
/note="assembly fragment:02522
fragment_chain:1"
111999..126181
/note="assembly fragment:01840
fragment_chain:1"
126282..159596
/note="assembly fragment:01914
fragment_chain:1
clone_end:SP6
vector_side:right"
BASE COUNT 54172 a 29956 c 28660 g 46068 t 700 others
ORIGIN

alignment_scores:
Quality: 407.50 Length: 447
Ratio: 1.827 Gaps: 19
Percent Similarity: 49.888 Percent Identity: 30.649

alignment_block:
US-09-775-181-2 x AL161654/rev ..
Align seg 1/1 to reverse of: AL161654 from: 1 to: 159596
372 ArgGlyProAspGlnHlsSerGlySerThrLysAspValSerGluCl 388
|||||
102605 AGGGTCCGGATCAGCATATTTAGGAAGTACAAAGATGTGTCAAGA 102556
|||||
400 uAlaTyrValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
|||||
102555 AGCCTATGCTGCTTACCTTGCAGGAGGGTGCCTTCTGTGCTGATG 102506
|||||
```

```

405 spSerProCysPheValGlnGluAspLysTyrLeuArgLeuAlaIle 421
|||||
102505 ACAGCCCATCTTGGTGCAGAAAGATTAATTTACGACTTGGCATCATC 102456
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
102455 TCCTTCCAGCCCTGTGTATGCTGCTGCATCTCGTTAGCATGCTGGTGGT 102406
438 lTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuL 455
|||||
102405 CTACCACTTTCGCAAGCAAG..... 102384
455 euGluThrIleLeuPheGlySerLeuLeuTyrPheProValValIle 471
|||||
102383 .....GTAAACCCAGGAACCTG.....GTTATGATC 102357
472 LeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaAr 488
|||||
102356 CTGTATTAC..... 102348
488 gLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisA 505
102347 .....A 102347
505 rgValLeuLys.ValPheLeuSerArgThrAlaGlnArgIleProTyrMe 521
|||||
102346 GAGCAACCTCTTATCTTTTAAAGCAAA.....CAGAAATTTCCAATTCA 102303
521 tThrGlyGlyArgValMetArgMetLeuAlaValIle..... 533
102302 AGTGTCTTCCTCCTTAGCAGCAGCTCTTCTACAATATGATGCGTAGA 102253
534 ..LeuLeuValValPheTrpPhe..... 540
102252 ACATGAAGTTATATTCTTTAGCTGACCAGAAAATAGATTAACAACAT 102203
541 .....LeuIleGlyTrpTh 545
102202 GATPACTTTATTTTAACATTTACAGACTGTTTCCCTTGTTTAGCCCTGGAT 102153
545 rSerSerValCysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnG 562
|||||
102152 CTCAGTGCT.....T 102142
562 lLysThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
02141 CATTTGACTTAAATGACTCTCTTTACATCATGCTTTGGCCTCTAGAGAA 102092
579 AspTyrMetThrAlaValAlaGluPheLeuLeuLeuTrpGlyValTy 595
|||||
102091 ACTTATTTCTACCAGCATCCAATGATATTATTTTCTATCCAAGAA 102042
595 rLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProArgT 612
102041 GCCTTGTAT..... 102032
612 yrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePheHis 628
102031 .....GTTTATTAT 102023
629 ThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrp...MetLe 644
|||||
102022 ATTTTCTTTCTTCTTAGTGTGTTTAAATA.....TGGTAAATGAC 101983
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
101982 TCACATATTTTACGCCATATATCTTTGAAGATGAATTTATTTATGACC 101933
661 euLeuLeu...lleProLysPheSerHisSerAsnProArgAsp 676
|||||
101932 TTATATTTTGTCTTACATTAGTGTCCAGCAATAATAACAATCCCGGAAG 101883

```

```

677 AspIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGl 693
|||||
101882 AACTTA.....CACAGTTTGGAGAGAACATGCTCTTTGGTCTAGTCT 101839
693 ySer.TyrLeuAsnSerSerIleAsnSerAlaTrp..... 704
|||||
101838 AGATANTCAAGAGTAGTACAATCCAGAGCAGTGGCTCATGCGCTGAATC 101789
705 SerGluHis.....SerLe 709
|||||
101788 TCAGCACAATTTGGGAGCCCAAGGCTGGAGAATTCCTTGAGGCCAGAGATT 101739
709 u.AspProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeu 725
|||||
101738 GAGACCAAGCTGACCAACATAGCAGACCGCTGCTCTACAAAAAACA 101689
726 GluIleTyrLysArgLysLysMetIleThrAsnAsn.ProHisLeuGlnL 742
|||||
101688 CAAATTTTAAAGCTAGCCAAATGTGGTGGTCATGCTATAGTCCCA 101639
742 yLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgArgIle 758
|||||
101638 CTATTTAGGAGGCTGAGAAAGGAGTA.....TCTCTTGAGTCTAGACTT 101595
759 ThrGluIleProGluThrValSerArgGlnCys 769
|||||
101594 CAAGACTAGCTAGCTATGATCACACCACTGT 101562
seq_name: gb_htg5:AC017271
seq_documentation_block:
LOCUS AC017271 10892 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017271
VERSION AC017271.1 GI:5553715
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 10892)
ADAMSM Adams,M. and Venter,J.C.
REFERENCE Direct Submission
AUTHORS Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210111 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source Location/Qualifiers
1..10892
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3383 a 2213 c 2125 g 3171 t
ORIGIN

```

```

alignment_scores:
Quality: 338.50 Length: 566
Ratio: 1.273 Gaps: 25
Percent Similarity: 46.996 Percent Identity: 23.675

```

```

alignment_block:
US-09-775-181-2 x AC017271/rev ..

```

```

Align seg 1/1 to reverse of: AC017271 from: 1 to: 10892

```

```

226 LeuArgArgLysTrpArg...Pro.HisLeuHisArgGlyProAsnG 241
|||||

```

488 ArgLeuLeuGlyIlePheAlaThrValTyrGlyThrValThrValThrLeuLysLeuHi 504
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8726 CGGGAGCTGGCGTCATCACCACGTCTCAAGGCAGCCATCATCATGAAGTTGTA 8677

504 sArgValLeuLysValPheLeuSerArgTrhAlaGlnArgIleProTyrM 521
:||| |||| |||| |||| ||||| ||||| ||||| ||||| ||||| |||||
8676 CCGCCACCTGGTGACATTTCGCCACCCGGAAGGCATCGCTGG...GTGC 8630

521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuValVal 537
::: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| |||||
8629 TCGCGCAGCTGGACCTGCTCAAGTATCTGGGCACCATGGTCTTTGGCGTC 8580

538 PheIrppHeLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGluLy 554
::: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| |||||
8579 ATCTGCTACATGGCGCGCTTACGGCTCGTGGTGGATCTTCGGAAG 8530

554 s...GlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisLeuIleP 570
:: ||||| ::||| ::||| ::||| ||||| ||||| ||||| ||||| |||||
8529 TCGCACGTGGAGCGCTCAGGGAGCGGCACAG..... 8496

570 heAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaValAlaGlu 586
||| |||| |::||| ::||| ::||| ::||| ::||| ::||| |||||
8495 .AAACACTGCATCCCGCTCAAGTGGGAACGTGTCTACGCAGACACCGAG 8448

587 PheLeuPheLeuTtpGlyValTyrLeuCystyrAlaValAvalGrhVra 603
::||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| |||||
8447 ATGCTCATCTCTGCTTTGGACTGCACCTGTCCATCCCACTCGGAATGC 8398

603 lProSerAlaPhe..... 607
::: |||
8397 CAACACCAGTTCCGGGTAAAGATACAGTCAAATCGATTTCGACATCTTT 8348

607 607

8347 CTTTATCAATTAAAGCGCAAGAGATTGCTTAGATGATTAACTAAGTAGAA 8298

607 607

8297 GGGGAATTTAAGTCAAGTCAAAATGAAATTAATCGGTAATTTCTTTTGTCC 8248

608His_GluProArgTyrMetalAlaValAlaValHisAsnGluLeuIle 622
::: |||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| |||||
8247 CATTTGAAGAAGACCAATTTCTGTCTACCGCAGTCAGCGCTGGAGTTCCGT 8198

623 lIleSerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuGl 639
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| |||||
8197 GTCTCGTCCAGCTTCTACTTTCCTGGCTTTGCTCTACCTCCGGAAATGAG 8148

639 nSerAspTrpMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValT 656
::: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| |||||
8147 TCCCAGCGCCCATCTGTGTGGCCCTGTTTCATCCGCTCCACGCTGACGAATA 8098

656 hrValThrIleGlyLeuLeuLeuIleProLys.....PheSerHisSer 670
:: ||||| ||||| ::||| ::||| ||||| ||||| ||||| ||||| |||||
8097 GTTTCGCCCTGGGTCTGATATTGTCGCAAGATTGCTGTTACCAGCACAG 8048

671 SerAsnAsnProArgAspAspileAlaThrGluAlaTyrGlu 684

8047 CAGGTAAACC.....GATTTCAGTTTCGAACAGCATATGAT 8015

seq_documentation_block:			
LOCUS	AC009845	183766 bp	DNA
DEFINITION	Drosophila melanogaster, chromosome 2L, region 25A-25B, BAC clone BACR04C19, complete sequence.		INV 24-FEB-2001
ACCESSION	AC009845		
VERSION	AC009845.7	GI:13122706	
KEYWORDS	HTG.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;		

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 183766)

REFERENCE
AUTHORS
Celnikier, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, J., An, H., Baldwin, J., Beeson, K.Y., Beeson, R.A., Busam, D.A.,
Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fierriera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Hock, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, J.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacile, J., Parag, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirkas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 2L, region 25A-25B
Unpublished

TITLE
JOURNAL

2 (bases 1 to 183766)

REFERENCE
AUTHORS
Celnikier, S.E., Agbayan, A., Arcaluna, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacile, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

TITLE
JOURNAL

Submitted (03-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 24, 2001 this sequence version replaced gi:6532035.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES
source

Location/Qualifiers
1..183766
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="25A-25B"
/clone="BACR04C19 (D1087)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6)"

BASE COUNT 52592 a 39362 c 39573 g 52239 t

ORIGIN

alignment_scores:

Quality: 338.50 Length: 566
Ratio: 1.273 Gaps: 25
Percent Similarity: 46.996 Percent Identity: 23.675

alignment_block:

US-09-775-181-2 x AC009845/rev ..

Align seg 1/1 to reverse of: AC009845 from: 1 to: 183766

226 LeuArgArgLysTrpArg...Pro.HisLeuHisArgArgGlyProAsnG 241

11376 CTTTCAGAGTTGCGCGGTGCTTCATTGCGCGAGAGGATGTGTC 11327

241 lnglyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLeuGly 257

11326 AA.....CGATGGCTGGAGGAAGTTTTCGG..... 11301
258 GlyAspLysSerHisPheLysTrpSerProTyrLeuGluCysGluAs 274
11300TCGCGCTCATGGGTGAGTTGGCCCA 11275
274 nGlySerTyrLysProGly.....TriLeuValThrLeuSerS 287
11274' AACTTTATACGATAATTCGTGACATTAGTTCGGATAATA..... 11235
287 erAlaIleTyrGlyLeuGlnProAsnLeuValProGluPheArgGlyVal 303
11234TATATAATACCATATCTC..... 11217
304 MetLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSe 320
11216TGTAGTGTG 11209
320 rAspGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluC 337
11208 CGAT.....CGAATACGACCTTCTGCTGCTCACCAGACAAC 11168
337 ysMetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIle 353
11167 CGCGCCGCCACCCCGGATGTG.....TACACCTGTCTG 11136
354 CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnAsnPheAr 370
11135 TGCCGGGAATCTTACTACCTGCCCAAT...TCCAGCTCCAGGATTCGG 11089
370 gArgGlyProAspGlnHisIleSerGlySerThrLysAspValSerG 387
11088 T.....GGCGATCGAGTGGAACTGTCCG 11066
387 lu.....GluAlaTyrValCysLeuProCysArgGluGlyCysProPhe 401
11065 AGGGTCATCGACAACTACTCGTTCGTTCCGCGCGGATGCACATAAC 11016
402 CysAlaAspAspSerProCys.....PheValGlnGlu..... 412
11015 TCGGATACGAACCGCGTCTGCTGACCTTTTCAGGAGGAGGAGGTCTCAA 10966
413AspLysTyrLeuArgLeuAlaIleIleSerPheGlnGlyLeuCysM 428
10965 TGTGCACGCTCTCTCGCCCTCTCTGGTGGCCATCTGCTACTGGCGCTGCA 10916
428 eLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysAla 444
10915 TCCTGTGTGCATTCCTCGGGGTGATTGTCTCCGCGAGCGAAAGTGC 10866
445 LysSerIleArg..... 448
10865 AAGGTATTAGTTTACATATATCTAAAGTTGGAAACTTTGAGTAATCCTT 10816
449AlaSerGlyLeu...IleLeuLeuGluThrI 458
10815 TCATCATTTTCAGGCCATTGCTGTGATGTGGACTGTCTGGAGACGA 10766
458 leLeuPheGlySerLeuLeuLeuTyrPheProVal.ValIleLeu..... 472
10765 TACTCTGGGCATTTGTTTACTTTATGATCTGTGAGTGTCTCTGGCGAT 10716
472 472
10715 GGTCACTTCAACTTCTGTGTATTTACATATATTCATCTTAGTTGCGG 10666
473TyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAla 487
10665 TCCATTTCTTCCGCATCCACCGAGCGCTGCTTCTGGAGCCCTGGCTT 10616
488 ArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHi 504
10615 CGGGAGCTGGGCTTCATCACCTGTCTACGGCGCCATCATACTACTGAAAGTGTGA 10566


```
504 sArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyrM 521
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10565 CGCCACCTGTGTGACATTTGCGACCCCGGAAGCGCATCGTGG...GTGC 10519
521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuValVal 537
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10518 TGGCGAGCTGGACCTGCTCAAGTATCTGGCCACCATGGCTTTGGCGTC 10469
538 PheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGluLy 554
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10468 ATCTGCTACATGGCGCTTTACGGCCTCGTGGTGGATCTTCGGAAG 10419
554 s...GlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisLeuIleP 570
: ||| ||| ||| : : : : : |||
10418 TGGCGACCTGGAGACCTCAGGAGGCGGACAGC..... 10385
570 heAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaValAlaGlu 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10384 ..AACACCTGCCATCCCTCAAGTGGGAAGTGGTACGACGACGACGAG 10337
587 PheLeuPheLeuLeuTrpGlyValTyrLeuCysTyrAlaValArgThrVa 603
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10336 ATGCTCATCTGCTGTTTGGACTGCACCTGTCCATCCCGAGTGGGAATGC 10287
603 lProSerAlaPhe..... 607
10286 CAACACCCAGCTTCGGGTAAGATGTACAGTCAAAATCGATTGCAGATCTTT 10237
607 ..... 607
10236 CTTTATCAATTAAGCGCAAGAGATGGCTTAGATGATTAAGTAGAA 10187
607 ..... 607
10186 GGGGAATTTAAGTCAAGTCAAAATCAATTAATCGTAAATTTTCCTTTTGTCC 10137
608 ...His.GluProArgTyrMetAlaValAlaValHisAsnGluLeuIle 622
: : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10136 CATTGAAGGAACGACAAATTTCTGGTACCACGACTGACGCTGGATTCCTG 10087
623 lLeSerAlaIlePheHisThrIleArgPheValLeuAlaSerArgLeuGl 639
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10086 GTCTCGTCCAGCTCTACTTCTGCGCTTTGTCTACCTGCCGGAATGAG 10037
639 nSerAspTrpMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValT 656
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10036 TCCAGCGCCATCTGTGCGCCCTGTTTCATCCGCTCCCGACGTGACGAATA 9987
656 hrValThrIleGlyLeuLeuLeuIleProLys.....PheSerHisSer 670
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9986 GTTTCGCTTGGGTCTGATATTTGTGCCAAAGTTGTGGTACCACGACAAAG 9937
671 SerAsnAsnProArgAspAspIleAlaThrGluAlaTyrGlu 684
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9936 CAGTAACCC.....GATTTCAGTTCGAAAGCATATGAT 9904
```

seq_name: gb_in1:AE003575

```
seq documentation block:
LOCUS AE003575 327209 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386046 section 16
of 16, complete sequence.
ACCESSION AE003575 AE002638
VERSION AE003575.2 GI:10727324
KEYWORDS HTG
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 327209)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
```

Amantides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Basu,A., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garq,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,R.A., Zhong,W., Zhou,X., Zhu,S., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of *Drosophila melanogaster* Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 327209)

Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.

Direct Submission

Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

On Oct 9, 2000 this sequence version replaced gi:7295650.

Location/Qualifiers

1..327209

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="2L"

join(<29815..29907,30773..>30955)

/gene="CG11933"

/product="CG35796"

/db_xref="FLYBASE:FBan0011933"

/db_xref="FLYBASE:FBgn0031615"

/evidence=not_experimental

<29815..>30955

/gene="CG11933"

/map="25A3-25A3"

/db_xref="FLYBASE:FBan0011933"

/db_xref="FLYBASE:FBgn0031615"

/evidence=not_experimental

join(29815..29907,30773..30955)

/gene="CG11933"

/note="CG11933 gene product"

/codon_start=1

/db_xref="FLYBASE:FBan0011933"

/db_xref="FLYBASE:FBgn0031615"

/evidence=not_experimental

/protein_id="AAF50997.1"

/db_xref="GI:7295691"

/translation="MAQPLGSPDCHLWTGELGGSGFCGSGCDYLRRHILSPNPDQ"

364 euProValAsnAsnPheArgArgGlyProAspGlnHisLeuSerGly 380
1387AAAACCTCTCCCAAG..... 1401
381 SerThrLysAspValSerGluGluAlaTyrValCysLeuProCysArgG1 397
1401 1401
397 uGlyCysProPheCysAlaAspAspSerProCysPheValGlnGluAspL 414
1402GATCGCACTTTGATCTACATCGACGACATC 1432
414 ys.....TyrLeuArgLeuAlaIleLeuSerPheGlnGly 425
1433 AGGTCAATCCAAACCATATATATGTCGGCTAGTGTTCGTCATTGGA 1482
426 LeuCysMetLeuLeuAspPheValSerMetLeuValValTyrHisPheAr 442
1483 GTGATTATTGCCACAGTTTTCTGGCCTTAAACATTAAAGTAT.....CG 1526
442 gLysAlaLysSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleL 459
1527 CAATCAAGATACATCAAGATGTCCAGTCCCATTTGAACAATCTGATCA 1576
459 euPheGlySerLeuLeuLeuTyrPheProValValIleLeuTyrPheGlu 475
1577 TTGGGGCTGTATGATGACCTATTGAGCATCATTTTCTGGGTCTCGAT 1626
476 ProSer.....ThrPheArgCysIleLeuLeuLeuAr 485
1627 ACCAATTAAGTAGTGGCAGCTTTTCCCTATATCTGCACAGCTCGAGC 1676
485 gTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuL 502
1677 CTGGATCTTGATGGCTGATCAGTCTCAGTTTGGAGCCATGTTCTCGA 1726
502 ySLeuHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIle 518
1727 AGCGTGGCGGGTGCTGATGATATACCGCATCTGAAGCTCAATAG.... 1773
519 ProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLe 535
1774 AAGGTGATCAAGGACTCAATTTATGTTGTTGGCGCTGCTTTGGC 1823
535 uValValPheTrpPheLeuIleGlyTyr...ThrSerSerValCysGlnA 551
1824 CATGTATATAGCCATTATACCACTGGCAGATTGCCGATCCCTTTTACC 1873
551 snLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHis 567
1874 CGGAACCTAAACAGTTGGAACTTGTGATCAGCAGAGATATTGATGATGC 1923
568 LeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaVa 584
1924 TTGGTGATCCCGAAACGAGTACTGCCAGTCTGAGCACATGACCATATT 1973
584 lAlaGluPheLeuPhe.....LeuLeuTrpGlyValT 595
1974 CGTTAGCATATTATGCTCTACAAGGAGTCTGTTGTTGTTTGGCGCT 2023
595 yrLeuCysTyrAlaValArgThrValPro...SerAlaPheHisGluPro 610
2024 TTTTGGCTGGAACTCGACATGTTTCTATACCGGCTCTGAACGATTC 2073
611 ArgTyrMetAlaValAlaValHisAsnGluLeuIleLeuSerAlaIlePh 627
2074 AAGCATATTGTTTCTCGCTTAAACGTGTTTCATCCTTCTCTGGCGG 2123
627 eHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetL 644
2124 AGCGGTATATCCCTGGTGTCTCGGATCGAAGGATTTAGTTTTGTCT 2173
644 euMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGly 660

2174 TACTCTCGTTTTTT.....ATCATTTTTTACGACGACACTTTGTGT 2217
661 LeuLeuLeuIleProLysPheSerHisSerSerAsnAsnProArgAspAs 677
2218 TTGGTGTTCGTACCGAAATTTGTGGAGCTGAAGCGGAATCCCGAGGCGT 2267
677 pIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGlyS 694
2268 GGTGGCAAAACGCTTTAGGGCCAGCTTGAGACCCATGTCCAAAACGAC 2317
694 erTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHisSerLeuAsp 710
2318 GC..... 2319
711 ProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluI 727
2320CGGGATTCC.....TCGGTGTCCGAATC 2342
727 eTyrLysArgLysLysMetIleThrAsn...AsnProHisLeuGlnLysL 743
2343 GGAGCAACGATTTGGAGATGTAAAGAACACAACTGCCGATTCGGAAGG 2392
743 ysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgIleThr 759
2393 CGCTGATGGAGAGGAGAACGAGCTGCAGGCTTAAATCCGCAAGCTGGGA 2442
760 GluIleProGluThrValSerArgGln.....CysSe 770
2443CCCGAG.....GCACGCAATGATCGATGGGTGACCTGCAC 2480
770 rLysGluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaL 787
2481 AGGTGGCTCCAACTCGGTAGCAACTGGAGCCCATACTGAACGATGACA 2530
787 euIleArgLysAsnProGluSerSerGlyAsnThrGlyLysSerLys 803
2531 TTGTTAGGCTCTCAGCTCCACCGGTGCTCGAGAGATGCCAGCACCA 2580
804 GluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerTh 820
2581 GAAGTTACCGAGATCAGCTCGCTGGATGTGTGACCTCGACTCATGTGGA 2630
820 rTyrAspHis.....ValArgAspGlnThrGluGluSerSerSerL 834
2631 GATGGATAACTCCTTTGTGCTGGTGCAGTCTACAGTGTATGCCCATGC 2680
834 euProThrGluSerClnGluGluThrThrGluAsn..... 846
2681 TTCCTCCCAAAAGAAAGCAATCGATTGTAGAGCACCACTCGCATGCC 2730
847SerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLe 861
2731 CCTGCTCCAATATGATCCACCCCTCCAGCAGCACTGACAGCAGCACTT 2780
861 uLysGluAspSerGluAlaGluSerThrGluSerValProLeuValCysL 878
2781 ACAGCAACATCAGCAGATGTCAGCAGCAGCAC.....CTGC 2815
878 ysSerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisPro 894
2816 AGCAGCAGCAACACCGACAGATGCAACAGCAACAGCAGCAGCAGCAT 2865
895 ArgThrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysG1 911
2866 CATCATCGCCCATCTGGAGAGAGAACTCGGTGTCCGCTCAGACCGCATGA 2915
911 uLysThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGly..... 924
2916 T.....AATATAGCAGCATCACCAGTACGGGGGCAAGCGGAGCG 2956
925ValGluGluArgThrLys..... 930

221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
:::|||||
901TATCAGGGCAAGACTCGCTGGCCAGCGAGCGCTGGCCAGCT 944
:::|
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
|||||
945 CCTCCCTGATCGCCATGCCCTCAGTACTTCCACGTTGGTGGCGGACCAATT 994
|||
250 rPArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
|||
995 GCATTGCTCTGAAGGC.....TGGGCA 1017
|||
267 ProProTyrLeuGlu.....Cy 272
|||
1018 GATCCCAAGGCTTCGGGAATTCCTGAAGAAGGTCCATCCCAAGGAGTCTG 1067
|||
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
|||
1068 TCCACAATGGTTTGGCAAGAGGTTTGGGAAGAACAATTAACTGCCAC 1117
|||
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
|||
1118 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGCACACCTTTCTCGAGAGG 1167
|||
295AsnLeuValProGluPheArgG 302
|||
1168 TCACGAAGAAGTGGGCACAGGTTTACACACAGCTCGACACGCTTCGGAC 1217
|||
302 LyValMetLysValAspIleAsnLeuGlnLysVal..... 313
|||
1218 CCCCTGTACAGGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1267
|||
313 313
1268 GATTACAGCGCATTTACGGATATCTACAATGTGTACTTAGCAGTCTACTC 1317
|||
314AspIleAspGlnCys..... 318
|||
1318 CATTCGCCACGCTTGAAGATATATATACCTGCTTACCTGGGAGAGGC 1367
|||
319SerSerAspGlyTrpPhe 324
|||
1368 TCTTCACCAATGGCTCTGTGTCACACATCAAGAAGTTGAGGCGTGGCAG 1417
|||
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
|||
1418 GTCCTGAAGCACCTACGGCACTAAACTTTACAAACAATATGGGGAGCA 1467
|||
336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
|||
1468 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGAACATATTCCATCATCA 1517
|||
348Gly 348
|||
1518 ACTGGCACCTTCCCCAGAGATGGCTCCATCGTGTGTTAAGGAAGTCGGG 1567
|||
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
|||
1568 TATTACACGCTTATGCCAAGAGGAGAAAGACTCTTTCATCAACAGGCA 1617
|||
362 yValLeuProValAsnAsnProArgArgGlyProAspGlnHis.... 377
|||
1618 GAAATCTGTGGAGTGGTGTCTCCAGGAGGTTGCTTCCCAACTGCA 1667
|||
378IleSerGlySerThrLysaspValSerGluGlu..... 388
|||
1668 GCCGAGACTCGCTGGCAGGACCAAGGAGGATCATTTGAGGGGGAGCCC 1717
|||
389AlaTyrValCysLeuProCysArgGluGly..... 398
|||
1718 ACCTGCTGCTTTGAGTGTGTGGAGTGTCTTCGTATGGGAGTATAGTGATGA 1767
|||

399CysProPheCysAlaAspAsp..... 405
|||
1768 GACAGATGCCAGTGCCTGTAAACAAGTCCCGCATGACTTCTGTGTCCAATG 1817
|||
406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
|||
1818 AGAACACACCTCTCTGGAGGCCATTCCTGTCCTGTTATCTTGTAGTGGAGT 1867
|||
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
|||
1868 GACATAGAATCTATCATAGCCATCGCCTTTTCTGCTGGGCATCCTCGT 1917
|||
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
|||
1918 GACCTCTTTGTCACCTCATCTTCGTTCTGTACCGGACACACCCGTGG 1967
|||
447 leArgAlaSerGlyLeuIleLeuGluThrIleLeuPheGlySerLeu 463
|||
1968 TCAATCTCAGTAGGAGCTCTGCTATATCATCTCGGTGGTGGTATTTTC 2017
|||
464 LeuLeuTyrPheProValValIleIlePheGluProSerThrPheAr 480
|||
2018 CTCGGCTATGTGCTCCCTTTCACCTCATCGCCAAACCTACTACCATC 2067
|||
480 gCysIleLeuLeuArgTrpAlaArgLeuGlyPheAlaThrValTyrG 497
|||
2068 CTGCTACTCCAGCGCTCTAGTTGGCTCTCTCTGCTGCTGCTACT 2117
|||
497 lyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
|||
2118 CTGCTTTAGTGACCAAAACCAATCGTATTTGACGCACTCCTGGCTGGCAGC 2167
|||
510PheLeuSerArgThrAlaG1 516
|||
2168 AAGAAGAAGATCTGCACCCGGAAGCCAGATTCATGAGCGCTTGGGCCA 2217
|||
516 nArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValI 533
|||
2218 AGTGATCATAGCTCCCTCATTTCTGTAGTAGTACAGTAACACTAGTGTGA 2267
|||
533 leLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCys 549
|||
2268 CCTTGATCATC..... 2278
|||
550 GlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnLysThrSerAs 566
|||
2279ATGGAGCTCCCATGCCCATTTGTCTACCCGAGTATCAAGGA 2322
|||
566 pHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrA 583
|||
2323 A.....GTCTACCTTATCTGCAATACAGCAACCTGGGTAGTGGCC 2366
|||
583 laValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysTyr 598
|||
2367 CTGTTGGTTTACATGGACTCTCATGAGCTGTACTACTATGCTCTTC 2416
|||
599 AlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlaVa 615
|||
2417 AAGACCCGCAACGTGGCGCAACTTCAATGAGCGTAATATACATCGCCTT 2466
|||
615 laValAlaHisAsnGluIlellelleSerAlaIlePheHisThrIleArgp 632
|||
2467 CACCATGTACACTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2516
|||
632 heValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPhe 648
|||
2517 TT.....GGGAGCAACTACAAGATCATCATCTACCTACCTGC 2548
|||
649 AlaHisThrHisLeuThrValThrValThrIleGlyLeuLeuLeuIlePr 665
|||
2549 TTCCGGGTGAGCCTCAGTGTGACGCTGGCGCTGGGTGCTGCTGCTGCT 2598
|||
665 oLys.....PheSerHisSerSerAsnAsnProArgAspAspI 678
|||

```
|||||
2599 GAAGATGTACATCATTCATGCCAAACCTGAGAGAACGTCGCGAGTGCCT 2648
678 leAlaThrGluAlaTyrgluAspGluLeuAspMetGly..... 690
2649 TCAGGACC.....TCGTGATGTTGTCGCGATGCACGTGCGTGATGGC 2689
691 .....ArgSerGlySerTyrgluAsnSerSerIleAsnSerAl 703
2690 AAAGTCGCGTGCCTCAACACCTTCTCTCAAC..... 2722
703 aTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysL 720
2722 ..... 2722
720 ysLeuTyraGlnLeuGluIleTyrgluAspGluLysMetIleThrAsn 736
2723 .....ATTTCGGAGAAAGAACCCGGGGCAGGG 2752
737 AsnProHisLeuGlnLysLysArgCysSer..... 746
2753 AATGCCAAATTTCTAACGGCAAGTCTGTGTCATGCTGAACACAGTGAAG 2802
747 LysLysGlyLeuGluArgSerIleMetArgArgIleThrGluIleProG 763
2803 ACAGGCGCCCAAGGAGACAGCAGTGTGGCAGCGCCTCTGTGCACGTGA 2852
763 luThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHis 779
2853 AGACCAACGAGAGCGCTGTAAACAA..... 2878
780 GlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProGluSerSe 796
2879 .....ACAGCGTGAATCAAA.....CCCCCTCACTAA 2904
796 rGlyAsnThrGlyLysSerLysGluThrLeuLysAsnArgValPheS 813
2905 AAGTTACCAAGGCTGTGGCAAGACCTGACCTTTTCAGAT..... 2944
813 erLeuLysLysSerHisSerThrTyraPheHisValArgAspGlnThrGlu 829
2944 ..... 2944
830 GluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluAs 846
2944 ..... 2944
846 nSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysG 863
2945 .....GCCAGCACCAAGACCCCTTTACAATGTGGAG 2975
863 luAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSer 879
2976 AAGAGACAAATACCCCTTCTGTCTACCTTCCAGCCCTCCAGCAGCCCTTCT 3025
880 AlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgTh 896
3026 ATGGTGTGCACCGCAGCGGCGCCACCGTGGCCACACACACCTCTGCC 3075
896 rSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysT 913
3076 ACCCATCTG.....ACCGCAGAGAGACGCC 3101
913 hrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThr 929
3102 CCTGTCTCTGCTGATTCGCTCATCCCAAGGCTG..... 3139
930 LysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSe 946
3140 .....CCTCCTCTCTCCGCGAGCAG..... 3163
946 rAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnP 963
```

```
3164 .....CCACAGAGCCGCCCC 3179
963 roAlaGluProArgLysProGlnLysSerGlyIleMetLysGlnGln 979
3180 CTCAGAGCCCCCGCAGAGCCCAAG.....TCCCTGATGGACCACTG 3223
980 ArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3224 CAAGGCGTAGTCACCAACTTCGGTTCGGGATTCAGATTTCCATCCGCT 3273
991 .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1005
3274 GCTGCAGGCCCGGGACACCA.....GGAA 3299
1005 luValCysProTrpGluValTyraPheLeuThrProGlyProValProSer 1021
3300 AC.....AGCCTGCGCTCTCTACCGCCCGCTCCGCCG 3337
1022 Glu.....SerLysValGlnLysHisValSerIleValAlaSerGluMe 1036
3338 CAACACCTGCAGATGCTGCCCTGCACCTGACACCTTCCAGGAGGATC 3387
1036 tGluLysAsnProThrPheSerLeuLysGluLysSerHisLysProL 1053
3388 CATCTCCCTCTCTGGGAGGACATCGATGACAGT..... 3424
1053 ysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAla 1069
3425 .....GAGAGATTCAGCTCCTG 3442
1070 GluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
3443 CAGGAGTTCGTGTAGCAGCGCGAAGGAAACACCGAAGAAGATGAATTGA 3492
1083 uAspGluLysLeuLeuIleSerLysThrProValLeuProGluArgAlaL 1100
3493 AGAGAGGAGGAGCTGCCACAGCAGCAAGCTGACCCCTGAGGATTC 3542
1100 ysGluGluAsnGlyGlyGlnProArgAlaAlaAsnValCysAlaGlyGln 1116
3543 CTGCTCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3592
1117 SerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluAs 1131
3593 TCAGTGCCCAAGTTCCTCCCTGATCTGAGTCGCTCTGCACCCCTCCAAA 3642
1131 nGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSer 1147
3643 TGTAACCTACGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3691
```

seq_name: gb_in2:AF318273

seq_documentation_block: 3663 bp mRNA INV 28-FEB-2001

LOCUS AF318273
DEFINITION Drosophila melanogaster metabotropic GABA-B receptor subtype 2 (GABA-B-R2) mRNA, complete cds.

ACCESSION AF318273

VERSION AF318273.1 GI:13160944

KEYWORDS

SOURCE fruit fly.

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

REFERENCE

AUTHORS

TITLE

397 uGlyCysProPheCysAlaAspSerProCysPheValGlnGluAspL 414
 1294TyrLeuArgLeuAlaIleIleSerPheGlnGly 425
 414 ys.....TyrLeuArgLeuAlaIleIleSerPheGlnGly 425
 1325 AGGTCAATCAACCATATATATGATCGGCTTAGTCTCGGTCAATTGGA 1374
 426 LeuCysMetLeuLeuAspPheValSerMetLeuValValTyrHisPheAr 442
 1375 GTGATTATGGCCACAGTTTCTGGCTTAAACATTAAAGTAT.....CG 1418
 442 glysAlaIleSerIleArgAlaSerGlyLeuLeuLeuGluThrIleL 459
 1419 CAATCAAGATACATCAAGATGTCCAGTCCCATTTGAACAATCTCATCA 1468
 459 eupheGlySerLeuLeuLeuTyrPheProValValIleLeuTyrPheGlu 475
 1469 TTGTGGGCTGTATGATTACCTATTGAGCATCATTTCTCGGCTCTCGAT 1518
 476 ProSer.....ThrPheArgCysIleLeuLeuAr 485
 1519 ACCACATTAAAGTAGTGGCAGCTTTTCCCTATATCTGCACAGCTCGAGC 1568
 485 gtrPAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuL 502
 1569 CTGGATCTTGATGGCTGGATTTCAGTCTCAGTTTTCGAGCCATGTTCTCGA 1618
 502 ysLeuHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIle 518
 1619 AGAGTGGCGGGTGCATTTCGATTATTCACCGATCTGAAGCTCAAG... 1665
 519 ProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLe 535
 1666 AGGTGATCAAGCACTATCAATTGTTTATGTTGGTGGGCGTCTTTGGC 1715
 535 uValValPheTrpPheLeuIleGlyTrp...ThrSerValCysGlnA 551
 1716 CATTGATATAGCCATTATAACCACTGGCAGATTGCGCATCCTTTTACC 1765
 551 snLeuGlnLysGlnIleSerLeuIleGlyGlnGlyThrSerAspHis 567
 1766 GCGAACTAAACAGATGGAACTTGCATCAGCAGAAATATTGATGATGC 1815
 568 LeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaVa 584
 1816 TTGGTGATCCCGAANACAGTACTGCCAGTCTGACCATGACCATATT 1865
 584 lAlaGluPheLeuPhe.....LeuLeuTrpGlyValT 595
 1866 CGTTAGCATATTATGCTTACAAAGGAGCTGTTGTTGGTTTTCGGCCT 1915
 595 yrLeuCysTyrAlaValArgThrValPro...SerAlaPheHisGluPro 610
 1916 TTTTGGCTGGGAACTGCACATGTTCTTATACCGGCTCTGAACGATCC 1965
 611 ArgTyrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePh 627
 1966 AGCATATTGGTTTCTCCGTTTAAACGTTTCACTCATCTGCTGGCGG 2015
 627 eHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetL 644
 2016 AGCGGTATATCCCTGGTCTATCGATCGAAAGGATTAGTTTGTGT 2065
 644 euMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGly 660
 2066 TACTCTGGTTTTT.....ATCATTTTGTACACACAGCCACTTGTGT 2109
 661 LeuLeuLeuIleProLysPheSerHisSerAsnProArgAspAs 677
 2110 TTGGTGTCTGACCCANATTGTTGGAGCTGAAGCGGAATCCCGAGGCGT 2159
 677 pLeaLaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSerGlyS 694
 2160 GTTGGCAAAACGGTTAGGGCCACGTTGAGACCCATGTCCAAAACGGAC 2209
 694 eTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHisSerLeuAsp 710
 2210 GC..... 2211
 711 ProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluL 727
 2212CGGATTCC.....TCGCTGTGCGAAT 2234
 727 eTyrLysArgLysLysMetIleThrAsnAsnProHisLeuGlnLysLysA 744
 2235 GGAGCAACGATTGCGAGATGTAAGAACAACA.....A 2266
 744 rGcysSerLysLysGlyLeuGlyArgSerIleMetArgIleThrGlu 760
 2267 ACTGCCGATTCCGA.....AAGGCGCTGATGGAGAAGGAGAACGAG 2307
 761 IleProGluThrValSerArgGlnCysSerLysGluAspLysGluGlyAl 777
 2308 CNG.....CAGGCTTAAATCCCAAGCTGGAGCCGACGCAAAATGGAT 2354
 777 aAspHisGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProG 794
 2355 CGATGGGTGACCTGCACAGT..... 2376
 794 luserSerGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArg 810
 2377GGTCCAAACGTGCTAGCAACTGGAGCCCATACTGAACGATGAC 2421
 811 ValPheSerLeuLys..... 816
 2422 ATTGTTAGGCTCTCAGCTCCACCGTGGCTCGAGAGATGCCCAGCACAC 2471
 817SerHisSerThrTyrAspHisValArg. 825
 2472 ACTTACCGAGATGACCTCGTGGATAGTGTGACCTCGATCATGTGGAGA 2521
 826AspGlnThrGluGluSerSerSerLeu 834
 2522 TGGATAACTCTTTCTGCTGGTGGCAGTCTACAGTGTGCGCCATCGCTT 2571
 835 ProThrGluSerGlnGluGluThrThrGluAsn..... 846
 2572 CTCTCCAAAAGAAAAGCAATCGATTGTAGAGCACCACTCCCATGCCCC 2621
 847SerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuL 862
 2622 TGCTCCAACTATGATGCGAGCCCATCCAGCAGCACTGCAGCAGCACTTAC 2671
 862 ysGluAspSerGluAlaGluSerThrGluSerValProLeuValCysLys 878
 2672 AGCAACATCAGCAGATGCGAGCAGCAC.....CTGCAG 2706
 879 SerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProAr 895
 2707 CAGCAGCAACACCAGCAGATGCAACAGCAACAGCAGCAGCAGCATCA 2756
 895 gThrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluL 912
 2757 TCATCGCCATCTGGAGAAAGAACTTCGCTGCTCAGACCGGATGAT. 2805
 912 ysThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGly..... 924
 2806AATAGGAGCAGCATCAGTACGGCGGCAAGCGAGCGGA 2847
 925ValGluArgThrLys..... 930
 2848 CGAGACTCTCCAGCATCGGAGAGGCGTCAATCGACCGCTCCAGCA 2897
 931SerGlnLysProLeuProLysAspLysGluThrAsnA 943

254AspGlyLeuGlyGlyAspLysSerHisPhe. 263
1338 CCTCATGAAGACCAACTTTACTGGGGTCTCTGGGGACATGATCTTGTTG 1387
264Lys 264
1388 ATGAGATGGTGACTGCCAGGAAGATATGAATCATGAATTTTAGAAA 1437
265 TrpSerProTyrLeuGluCysGluAsn...GlySerTyrLysProGly 280
1438 ATCGGGAAGGACTACTTTGATTATATCAATGTTGGCAGCTGGCCACAACGG 1487
280 TyrLeuValThrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuV 297
1488 TCAGCTGAAATGGATGATGATAAATATGTCAGAGAAAAATAATATCA 1537
297 al...ProGluPheArgGlyIleValMetLysValAsp 307
1538 TCAGATCTGTGTGAGTGAACCCCTGTGAAAAGGACAGATAAAGGTG... 1584
308 IleAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTyrPh 324
1585ATTGCTAAGAGAGAGTGAAGTGGTGT...TGG... 1614
324 eSerGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProIleL 341
1615ACATGCACCTCTGTA 1630
341 ysGlyLeuGlyPheValLeuGlyAlaTyrGluCys...IleCysLysAla 356
1631 AAGAGAATGAGTACGCTCTTTGATGAATACACGTGCAAGGCGCTGCCAGCTC 1680
357 GlyPheTyrHisProGlyValLeuProValAsnAsnPheArgArgGly 373
1681 GCCTCTCG... 1689
373 yProAspGlnHisIleSerGlySerThrLysAspValSerGluGluAla 390
1690 .CCCAATGATCAGCTTACAGT... 1710
390 yrValCysLeuProCysArgGluClyCysProPheCysAlaAspSer 406
1710 1710
407 ProCysPheValGlnGluAspLysTyrLeuArg... 417
1711 ...TGTGACCTCATCCAGTCGCGATCTCAGATGGGTGATCCTGAACC 1757
418 .LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValS 434
1758 AATTCACAGCACTTCTTTTTCGGTCCCTGGGTCTGCTGGCAACCTTGTTG 1807
434 erMetLeuValTyrHisPheArgLysAlaLysSerIleArgAlaSer 450
1808 TTACGGCTATATTCAATGTACGGCATACTCCAGTGGTCAAACTCTCC 1857
451 GlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTyrPh 467
1858 AGCCGAGAACCTTCTGCTACATCATCTAGCTGGCATCTGCTGGGTACTT 1907
467 eProValIleLeuTyrPheGluProSerThrPheArgCysIleLeuL 484
1908 GTGCAGCTTTCGTCTCATCCGCAAACTCAACAGATTTACTGCTATCTTC 1957
484 euArgCTTAlaArgLeuGlyPheAlaThrValTyrGlyThrValThr 500
1958 AACGAATGGCATCGGTCTCTCCCAAGCTATGATGATTATTCTGCTCTAGTA 2007
501 LeuLysLeuHisArgValLeuLysVal... 509
2008 ACTAAACCAACCCGATTCGCAAGAACTCTGGCTGGCAGCAGCAAGAAAT 2057
510PheLeuSerArgThrAlaGlnArgIleProT 520

3552 3552

```
1074 rpGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuLeuIleSer 1090
      ::|||  ::  |||:::
3553 .....CAAGACCTTCAGAGAGCTGGTAGCT 3576
      ::|||  ::  |||:::
1091 LysThrProValLeuProGluArgAlaLysGluGluAsnGlyGlyGlnPr 1107
      |||||  |||  ::  ::::|||::
3577 TTAACCTCCTCTCTCCTTTTAGAGACTCCATCGATTCTGGAAGCGCATC 3626
      :::::  |||  ::  ::::|||::
1107 oArgAlaAlaAsnVal.....CysAlaGlyGlnSerGluG 1119
      :::::  |||  |||  ::::|||  ;
3627 TCCTAGCTCTCCAGTTTCGAATCAGCTCTCTGTATCCCATCTTCCCCAA 3676
      ::  ::  ::  |||:::
1119 luLeuProProLysAlaValAlaSerLysThrGluAsnGluAsn 1133
      ::  ::  ::  |||:::
3677 AATAGCACACACTCCTCTCATCAGAGATTACACTCAAAGTTCTTCT 3720
```

OM of: US-09-775-181-2 to: N_Geneseq_0601:* out_format : pfs

Date: Nov 1, 2001 11:02 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+pn2n.model -DEV=xlh
-Q/cn2_1/USPTO.spool/US09775181/runat_30102001_161056_13847/app_query.fasta_1.1955
-DB=N_Geneseq_0601 -QEMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09775181_@GNI_1_509 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-775-181-2
Query length: 1215
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 192.710000
```

score_list:

Sequence	Strd Orig	Zscore	Escore Len	Documentation
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAA42817	+	580.00	764.27	1.3e-34
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAA41583	+	260.00	348.06	2.0e-11
/SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:AAT86166	+	254.00	305.44	4.0e-09
/SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:AAQ25815	+	243.00	287.84	4.6e-08
/SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:AAQ25812	+	231.00	274.15	2.7e-07
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AA231058	+	217.50	257.55	2.2e-06
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AA231058	+	214.00	253.64	3.7e-06
/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80419	+	213.00	250.87	5.3e-06
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AA231057	+	210.00	251.33	5.0e-06
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AA231083	+	210.00	247.36	8.3e-06
/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80420	+	208.00	243.98	1.3e-05
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV21511	+	204.50	233.96	4.6e-05
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ29423	+	204.00	235.69	3.7e-05
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ50616	+	200.50	238.06	2.7e-05
/SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:AAT96167	+	199.50	235.65	3.7e-05
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ50615	+	199.00	235.77	3.7e-05
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ50615	+	198.50	234.09	4.5e-05
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ89484	+	198.50	234.09	4.5e-05
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31065	+	198.00	230.60	7.1e-05
/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:AAT03887	+	195.00	229.04	8.7e-05
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31060	+	193.50	224.75	0.0002
/SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:AAQ50879	+	193.00	225.84	0.0001
/SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:AAT86168	+	192.00	225.71	0.0001
/SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:AAT61362	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV26964	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV26964	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ50855	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ50855	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAV82485	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ89298	+	189.50	220.51	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ51398	+	188.50	221.02	0.0002
/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80418	+	188.00	218.20	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ50922	+	187.50	219.69	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31049	+	187.50	219.69	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ50617	+	186.00	219.72	0.0003
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV26965	+	182.50	210.32	0.0010
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV26965	+	182.50	210.32	0.0010
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ25056	+	182.50	210.32	0.0010
/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAV82486	+	182.50	210.32	0.0010
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ89298	+	182.50	210.32	0.0010
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAZ95887	+	179.50	203.60	0.0023

```
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ76700 + 179.50 194.69 0.0071
/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80416 + 179.00 208.13 0.0013
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV52159 - 179.00 191.06 0.0113
/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAV27324 + 178.00 214.23 0.0006
/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ51400 + 177.50 207.97 0.0013
seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:AAA42817
```

seq_documentation_block:

```
ID AAA42817 standard; cDNA; 368 BP.
XX
AC AAA42817;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:1557.
```

```
XX
OS Homo sapiens.
XX
PN WO200021990-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-US24205.
XX
PR 15-OCT-1998; 98US-0104435.
XX
PA (GEM ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Weisberg D, Treacy M;
XX
WP1: 2000-317937/27.
XX
```

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (sESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders -

Claim 1; Page 487; 618pp; English.

AAA41261 to AAA43419 represent specifically claimed secreted expressed sequence tags (sESTs), isolated from human, mouse, xenopus and rat tissue sources. The sESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antidiabetic; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be used for gene therapy and in vaccines. The sESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the sESTs. Proteins encoded by the sESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders

CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 368 BP; 115 A; 78 C; 79 G; 96 T; 0 other;

alignment_scores:
Quality: 580.00 Length: 112
Ratio: 5.179 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-775-181-2 x AAA42817 ..
Align seg 1/1 to: AAA42817 from: 1 to: 368

626 IlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTr 642
|||||
24 ATATTCATACATAGATTGTTCTTGCTCAAGACTTCAGTCTGATG 73

642 pMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrI 659
|||||

74 GATGTTGATGCTGTATTTGCACATACTCATTTGACTGTGACAGTCACCA 123

659 leGlyLeuLeuLeuPheProLysPheSerHisSerAsnAsnProArq 675
|||||

124 TTGGTGTGCTTTGATTCAAGATTTTCACATCAAGCAATTAACCCACCA 173

676 AspAspIleAlaThrGluAlaTyrGluAspGluLeuAspMetGlyArgSe 692
|||||

174 GATGATATTGCTACAGACATATGAGGATGAGCTAGACATGGCGCGATC 223

692 rGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHisSerL 709
|||||

224 TGATCCTACCTGAACAGCAGATATCAATTCAGCCTGGAGTGACACAGCT 273

709 euAspProGluAspIleArgAspGluLeuLysLysLeuTyrAlaGlnLeu 725
|||||

274 TGGATCCAGGACATCTCGGACGAGCTGAAAACTCTATGCCCACTG 323

726 GluIleTyrLysArgLysLysMetIleThrAsnAsn 737
|||||

324 GAATATATAAAGAAAGAGATGATCATCAAAACAAC 359

seq_name: /SIDSL/gcgdata/geneseq/NA2000.DAT:AAA41583

seq_documentation_block:

ID AAA41583 standard; cDNA; 182 BP.

XX
AC AAA41583;

XX
DT 21-AUG-2000 (first entry)

XX
DE Human secreted expressed sequence tag SEQ ID NO:323.

XX Human: mouse; xenopus; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; cytostatic; vulnery; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; neurotropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune diabetes; asthma; myeloid cell deficiency; ulcer;
KW insulin dependent diabetes; multiple sclerosis; allergic condition;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.

XX

OS Homo sapiens.

XX WO200021990-A1.

XX PD 20-APR-2000.

XX PF 15-OCT-1999; 99WO-US24205.

XX PR 15-OCT-1998; 98US-0104435.

XX PA (GEMY) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

XX PT Merberg D, Treacy M;

XX DR WPI: 2000-317937/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 243; 618pp; English.

XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective;
CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.

XX SQ Sequence 182 BP; 46 A; 36 C; 49 G; 51 T; 0 other;

alignment_scores:

Quality: 260.00 Length: 51
Ratio: 5.200 Gaps: 0
Percent Similarity: 98.039 Percent Identity: 98.039

alignment_block:

US-09-775-181-2 x AAA41583 ..

Align seg 1/1 to: AAA41583 from: 1 to: 182

518 IleProTyrMetThrGlyArgValMetArgMetLeuAlaValIleLe 534
|||||

27 ATTCATATATGACTGGCGGAGGGTTCATGAGGATCTGGCAGTAATACT 76

534 uLeuValValPheTrpPheLeuIleGlyTrpThrSerValCysGlnA 551
|||||

77 CTGTGAGTGTGTTTGGTCTTCATGGCTGGACTTCATCTGTGTGCCAGA 126

551 snLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHis 567
|||||

127 ATTTGGAGAAACAGATTTTCACATTTATTGGCCAGGGGAAAAACCCGATCAC 176

568 Leu 568

177 CTC 179

seq_name: /SIDS1/gcdata/geneseq/geneseqn/NA1997.DAT:AAT86166

seq_documentation_block:

ID AAT86166 standard; DNA: 4000 BP.

XX AC AAT86166;

XX DT 02-DEC-1997 (first entry)

XX DE Nucleotide sequence of pCar/R1.

XX KW Chimeric receptor; extracellular domain; seven transmembrane domain;
intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
MGR; calcium receptor; CR; mglur; G protein-coupled receptor;
neurological disease; ss.

XX OS Chimeric - Homo sapiens.

XX FH Key Location/Qualifiers
FT CDS 41..3700
FT FT /*tag= a
FT FT /product= pCar/R1

XX PN W09705252-A2.

XX PD 13-FEB-1997.

XX PF 25-JUL-1996; 96WO-US12336.

XX PR 26-JUL-1995; 95US-0001526.

XX PA (NPSP-) NPS PHARM INC.

XX PI Fuller FH, Hammerland LG, Krapcho KJ;

XX DR WPI; 1997-145690/13.

XX DR P-PSDB; AAW25763.

XX PT Chimeric receptors comprising metabotropic glutamate receptor and
calcium receptor - used for screening for neurologically active
compounds

XX PS Example 3; Fig 3; 177pp; English.

XX CC This sequence encodes a chimeric receptor. The chimeric receptor
comprises an extracellular domain, a seven transmembrane domain, and
an intracellular cytoplasmic tail domain, and a sequence of at least
6 contiguous amino acids is homologous to a sequence of a metabotropic
glutamate receptor (MGR), and a sequence of at least 6 contiguous amino
acids is homologous to a sequence of a calcium receptor (CR). The
chimeric receptor may be used for screening for compounds that bind to
or modulate the activity of MGR or CR and for determining the site-of-
action of a CR active compound. The compounds can be used in the
treatment of neurological diseases and disorders. They can also be
used as diagnostic agents. Chimeric receptors such as this, allow the
coupling of certain functional aspects of an MGR with certain functional
aspects of a CR. They allow for more efficient high-throughput screening
of compounds.

XX SQ Sequence 4000 BP; 955 A; 1128 C; 1010 G; 907 T; 0 other;

alignment_scores:

Quality: 254.00 Length: 1285
Ratio: 0.490 Gaps: 48
Percent Similarity: 40.311 Percent Identity: 18.599

alignment_block:

US-09-775-181-2 x AAT86166 ..

Align seg 1/1 to: AAT86166 from: 1 to: 4000

```

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspse 91
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 TCAGAGCACATTCCTCTACGATTGCTGTGGAGCAACTGGCTCAGG 483
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
-91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
484 CGTCTCCACGGCAGTGGCAATCTCTGGGGCTCTTCTACATT..... 526
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
527 .....CCCAGGTCAGTTATGCTCTCTCCAGC.....AGA 556
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
557 CTCCTCAGCAACAAGATCAATTCAAGTCTTCTCCGAACC.....ATCCC 603
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
604 CAATGATGAGCACCAGGCCACTGCGCATGGCAGACATCATCGATATTTC 653
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
158 ln.AlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
654 GCTGGAACTGGTGGGCACAATTGCGCTGATGACGACTATGGGGCGCG 703
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
173 .....ArgAlaAlaIleThrPheSerT 180
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
704 GGGATTGAGAAATTCGAGAGGAAGCTGAGGAAGGGATATCTGCATCGA 753
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
754 CTTCACTGAACATCTCTCCAGTACTCTGATGAGGAGAGATCCAGCATG 803
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
804 TGGTAGAGGTGATTCAAATTCACGGCCAAAGTCATCGTGGTTTCTCC 853
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
854 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATTGTCCGGCGCAA... 900
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
901.....TATCAGGGCAAGATCTGGCTGGCCAGCGAGGCTGGCCAGCT 944
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
945 CCTCCCTGATCGCCATGCCCTCAGTACTTCCACGTGGTGGCGGACCATT 994
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
250 rPArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
995 GGATTCGCTCTGAAGGC.....TGGGCA 1017
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ProProTyrLeuGlu.....Cy 272
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1018 GATCCAGGCTTCGGGAATTCCTGAAGAAGGTCCATCCAGGAAGTCTG 1067
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
-272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1068 TCCACAATGGTTTGGCAAGGAGTTTGGGAAGAACAATTTAACTGCCAC 1117
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
285 LeuSerSerAlaIleTyrGluGlnPro..... 294
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1118 CTCGAAGAAGGTGCAAAAGGACCTTTACCTGTGCACACCTTTCTGAGAG 1167
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....AsnLeuValProGluPheArgG 302
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1169 TCACCAAGAAAGTGGCAGACAGGTTTATAGCAACAGCTCGACACGCTTCCGAC 1217
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1267
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
313 ..... 313
1268 GATTACAGCATTTACGGATATCCTACAATGTGTACTTAGCAGTCTACTC 1317
314 .....AspileaspGlnCys..... 318
1318 CATTTGCCACGCCCTTGCAGAGATATATACCTGCTTACCTGGGAGAGGGC 1367
319 .....SerSerAspGlyTyrPhe 324
1368 TCTTCACCAATGGCTCCTGTGCAGACATCAAGAAAGTTGAGGGCTGCGAG 1417
325 SerGlyThrHisLysCysHisLeuAsnAnsSer..... 335
1418 GTCCTGAAGCACCTAGCGCATCTAAACTTTTACAAACAATATGGGGAGCA 1467
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1468 GGTGACCTTTGATGAGTGTGGTACCTGGTGGGGAACATATTCATCATCA 1517
348 .....Gly 348
1518 ACTGGCACCTCTCCCGAGGATGGCTCCATCGTGTTAAGGAAGTCGGG 1567
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGln 362
1568 TATTACAAGCTCTATGCCAAGAGGGAGAAAGACTCTTCATCAACGAGGA 1617
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
1618 GAAATCCCTGGAGTGTGGGTTCAGAGGAGGTGCCCTTCTCCAACATGCA 1667
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
1668 GCCGACACTGCCCTGGCAGGACCAGGAAGGATATTGAGGGGAGCCCC 1717
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
1718 ACCTGCTGCTTTGAGTGTGTGGAGTCTCTGTATGGGGAGTATAGTGATGA 1767
399 .....CysProPheCysAlaAsp..... 405
1768 GACAGATGCCAGTCCCTGTAAAGTGGCCAGATGACTTCTGTGTCCAATG 1817
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
1818 AGAACACACCTCTCGGAGCCCATTCCTGTCGGTTATCTTGAGTGGAGT 1867
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
1868 GACATAGAATCTATCATAGCATCGCTTTCTTGCTGGCATCCTCGT 1917
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
1918 GACCGCTTTTGTACCCCTCATCTTCCTGTACCGGACACACCCGTGG 1967
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
1968 TCAATCCTCCAGTAGGAGCTGTGTATATCATTTCTGGCTGGTATTTC 2017
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
2018 CTGGGTATGTGTGCTTTCACCTCATCGCCGCAACCTACTACCAATC 2067
480 gCysIleLeuLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2068 CTGCTACCTCCAGCGCTCCTAGTTGGCTCTCTCTGCGCATGTGCTACT 2117
497 lyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
2118 CTGCTTTAGTACCAAAACCAATCGTATTGACGCGCATCTCGCTGCGCAGC 2167
510 .....PheLeuSerArgThrAlaGln 516
2168 AAGAGAAGATCTGCACCGGAAGCCAGATTTCATGAGCGCTTGGGCCCA 2217
516 naArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValI 533
2218 AGTGATCATAGCTCCATTCCTGATTTAGTGTACAGCTAACACTAGTGGTGA 2267
533 leLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCys 549
2268 CCTGTATCATC..... 2278
550 GlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAs 566
2279 .....ATGAGCTCCCATGCCCATTTTGTCTACCCGAGTATCAAGGA 2322
566 pHISLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrA 593
2323 A.....GTCCTACCTTATCTGCAATACCAGCAACCTGGGTGTAGTGGCCC 2366
583 laValAla...GluPheLeuPheLeuTrpGlyValTyrLeuCysTyr 598
2367 CTGTGGTTACAATGAGCTCCTCATCATGAGCTGTACTACTATGCGCTTC 2416
599 AlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlaVa 615
2417 AAGACCGCAAGCTGCGCGCAACTTCAATGAGGCTAAATACATCGCCTT 2466
615 laValAlaHisAsnGluLeuIleIleSerAlaIlePheHisThrIleArgP 632
2467 CACCATGTACACTACCTGCATCTGGCTGGCTTTCTGTTCCCATTTACT 2516
632 heValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPhe 648
2517 TT.....GGAGCACTACAAGATCATCTACTACTCTGC 2548
649 AlaHisThrHisLeuThrValThrIleGlyLeuLeuLeuLeuLeuPr 665
2549 TTCGCGGTGAGCTCAGTGTACGCTGGCTGGCTGGTGTACTTACTCC 2598
665 oLys.....PheSerHisSerSerAsnAsnProArgAspAspI 678
2599 GAAGATGTACATCATTCATTGCCAAACCTGAGAGAACTCCGCGAGTGCCT 2648
678 leAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
2649 TCAGGACC.....TCTGATGTTGTCGCGCATGCACGTCGCTGATGCG 2689
691 .....ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAl 703
2690 AAACGTCCGCTGCCCTCCAAACACTTCCTCAAC..... 2722
703 aTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysL 720
2722 ..... 2722
720 ysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAsn 736
2723 .....ATTTCGGAGAAAGAACCCCGGCGAGG 2752
737 AsnProHisLeuGlnLysLysArgCysSer..... 746
2753 AATGCCAATTCCTAACGGCAAGTCTGTGTCATCTGTAACACAGGTGAAG 2802
747 LysLysGlyLeuGlyArgSerIleMetArgArgIleThrGluIleProG 763
2803 ACAGGCGCCCAAGGACACAGCAGTGTGGCAGCGCTCTCTGTGACGTGA 2852
763 luThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHis 779
2853 AGACCAACGAGACGGCTGTACACCA..... 2878
780 GlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProGluSerSe 796
```


CC GLU-G-R-mediated neuronal excitation and in diagnosis.
 XX Sequence 5236 BP; 1288 A; 1406 C; 1395 G; 1147 T; 0 other;
 alignment_scores:
 Quality: 243.00 Length: 1246
 Ratio: 0.489 Gaps: 54
 Percent Similarity: 39.888 Percent Identity: 19.904
 alignment_block:
 US-09-775-181-2 x AAQ25815 ..

Align seg 1/1 to: AAQ25815 from: 1 to: 5236

```

153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1103 GATCGCCCTGGCTCAGCTCTGTGGCCATTCAAGTCCAGAATCTTCGCC 1152
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 oSerIleSerArgAlaIleThrPheSerThrAspSerLeu..... 183
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1153 A.....GCTGTCGACATCCACAGATCGCTATTCTGCG 1187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrargGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1188 ACAAGCATAGACCTCAGTGACAAAACTTTGTACAAATACTTCTCCTGAGG 1237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 .....GluSerArgIle 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1238 GGTCCCTTCTGACACTTTCACGCCAAGGCCGATGCTCGACATAGTCAAGC 1287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1288 GTTACAACCTGGACCTATGCTCAGCAGTCCACACAGAGGAA..... 1330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1331 .....TTACGGCGAGAGTGGAAATGGATGCTTT 1357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1358 CAAGAAGCTGGCTGCCAGGAGGCT...CTGCATCGCACACTCGGACA 1404
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405 AAATCTACAGCAATGTCGCGAGAAGAGCTTTGACCGGCTCCTCGGTAAA 1454
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1455 CTCCGGGAGCGGCTTCCCAAGGC...CAGGGTTGGTCTGCTCTTGCGCA 1501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ArgLysAspGlyLeuGlyGlyAspLysSerHis...PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1502 GGGCATGACAGTCCGGGCTTACTGAGTCCCATGCCCGCCCTGGCGCTCG 1551
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....SerProProTyrLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 TGGGCGAGTCTCACTCATTTGAAGTATGATGGGCGACAGAGATGAA 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 GTCATCGAAGGCTATGAGTGGAGGCCAACCGAGGATCACAATAAAGCT 1651
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 TCAGTCTCCAGAGTCAGGTCAATTGATGACTACTTCTGAAAGTGAAGGC 1701
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1702 TGGACACCAACACAAAGGAATCCTTGGTTCCTGAGTCTTCTGCAACATGCG 1751
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu..... 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

1752 TTCCAGTGTGCGCTACCTGGACACACCTTTGGAAAAACCCCAACTTTAAGAA 1801
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1802 AGTGTGCACAGGAATGAAAGCTTGGAAAGAACTATGTCCAGGACACGA 1851
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1852 NAATGGATTGTTCATCAATGCCATCTATGCCATGGCACATGGCTGCAG 1901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 sLysCys...His.....LeuAsnAsnSerGluCysMe 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1902 AACATGCACCATGCTCTGTGCCCGCCATGTGGGCTGTGTGATGCTAT 1951
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1952 GAAACCCATGATGGCAGGAGCTCTCTGATTTCCTCATCAATCCTCTT 2001
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 ..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2002 TTGTCGGAGTGTCTGGAGAGGAGTGTGTTGTCGATGAGAAGGGGATGCT 2051
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 ...GlyAlaTyrGluCysIle..... 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2052 CCCGGAAGGTATGACATTGAATCTGCAGTACACAGAGCTAATCGCTA 2101
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 .....CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2102 TCATATGTCCACGTGGGACCTGCATGAGGAGTGTGATATATTGATG 2151
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 snPheArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2152 ATTCAAAATCCAG.....ATGAACAAAAGCGGAATGTAGCATCT 2192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 ValSerGluGlu..... 388
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2193 GTGTCAGTGCCTTTTAAAGGTCAGATTAAAGGTCATACGCAAGG 2242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2243 AGAAGTGAAGTCTGCTGGATCTGCACGGCTGCACAGAGATGAGTTG 2292
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 .....CysProPheCys..... 402
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2293 TGCAGGACGAGTTACCTGCAGAGCTGTGACCTGGGTGGTGGCCCAAC 2342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2343 GCAGAGCTCACAGGCTGTGAGCCCATTCCTGTCCGTTATCTGTAGTGGAG 2392
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2393 TCACATGAATCTATCATAGCCATGCCCTTTCTTGTCCCTGGCATCCTCG 2442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 euAspPheValSerMetLeuValTyrHisPheArgLysAlaLysSer 446
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2443 TGACGCTGTTGTACACCTCATCTCTGTTCTGTACCCGGACACACCCGCTG 2492
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLe 463
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2493 GTCAATCTCCAGTGGAGCTCTGCTATATCATTTCTGGCTGGTATTTT 2542
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 uLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2543 CTTGGCTATGTGTCCTTTTACCCCTCATCGCCCAACCTACTACCACAT 2592
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 rGlyIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2593 CTTGTACTCCAGCGCTCCTAGTTGGCTCTCTCTTCTGCCATGTGCTAC 2642
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 GlyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2643 TCTGTTTAGTGACCAAAACCAATCGTATTGCACGCTCTCTGGCTGGCAG 2692
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

510 .....PheLeuSerArgThrAlaG 516
2593 CAAGAAGAAGATCGCACCGGAGCCAGCATTCATGAGCGCTTGGGCC 2742
516 InArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
2743 AAGTGATCATAGCTCCATTCTGATTTAGTGTACAGTAACACTAGTGGTG 2792
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerValCy 549
2793 ACCTTGATCATC..... 2804
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerA 566
2805 .....ATGGAGCGCTCCCATGCGCATTTTGTCTACCCGAGTATCAAGG 2847
566 sPHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
2848 AA.....GTCTACCTTATCTGCAATACCAACACCTGGGTGTAGTGGCC 2891
583 AlaValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCystY 598
2892 CCTGTGGTTACAATGACCTCTCATCATGAGCTGTACTACTATGCGCTT 2941
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlav 615
2942 CAAGACCCGCAACGTCGCGCAACTTCAATGAGGCTAAATATACATCGCCT 2991
615 alAlaValHisAsnGluLeuIleLeuSerAlaIlePheHisThrIleArg 631
2992 TCACCATGTACACTACCTGCTATCATCTGGCTGGCTTTCGTTCCTCCATTAC 3041
632 PheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuIleP 648
3042 TTT.....GGGAGCAACTACAAGATCATCACTACTCG 3073
648 eAlaHisThrHisLeuThrValThrValThrIleGlyLeuLeuLeuIleP 665
3074 CTTGCGGCTGAGCTCAGTGTGACGGTGGCCCTGGGGTGCATGTTTACTC 3123
665 rOlyLys.....PheSerHisSerAsnAsnProArgAspAsp 677
3124 CGAGATGTATCATCATCTATGCGCAACCTCAGAGGAACGTCGCGAGTGCC 3173
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
3174 TTCAGGACCTCTGATGTGTC...GGCAGGAGCTCGGTGATGGCAACT 3220
691 .....ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpS 705
3221 GCGGTGCGCTCCCAACACCTTCCTCAAC..... 3248
705 erGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLysLeu 721
3248 ..... 3248
722 TyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnPr 738
3249 .....ATTTCCGGAGAAAGAGCCCGG..... 3271
738 oHisLeuGlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleM 755
3271 ..... 3271
755 etArgArgIleThrGluIleProGluThrValSerArgGlnCysSerLys 771
3272 .....GCGAGGAATGCCAAGAA 3289
772 GluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuI 788
3290 GAG.....GCAGCCAGAT 3303
788 eArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLysGluG 805
3304 TCTCGGCCAGCAGCAGCTGTCGTCGACATCGCAGCTTTGAAAACCC 3353
805 LuThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerThrTyr 821
3354 CCACACTGCAGTGAATGTTTCTAACGGCAAGTCTGTGTATGGTCTGAC 3403
822 Asp.....HisValArgAspGlnThrGluGluSerSerSerLe 834
3404 CAGGTGGAAGACAGGCGCCCAAGGAGCAGCACGTGTGGCAGCGCTCTCT 3453
834 uProThr.GluSerGln.GluGluGluThrThrGluAsnSerThrLeuG 850
3454 GTCACGTGAAGACCAACGAGCGGCTGTAAACCAACAGCCGTATACAA 3503
850 uSerLeu.....SerGlyLysLysLeuThr..... 858
3504 ACCCTCACTAAAGTTACCAAGCTCTGGCAAGAGCCTGACCTTTTCAG 3553
859 .....GlnLysLeuLysGluAspSerGluAlaGlu 868
3554 ATGCCAGCACCAAGACCTTTACAAATGTGAAGAGAGGACAATACCCCT 3603
869 SerThrGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLe 885
3604 TCTGCTCACTTCAGCCCTCCAGCAGCCCTCTATGTTGTGTGACCGCAG 3653
885 uSerSerGluLysLysThrGlyHisProArgThrSerMetLeuGlnLys 902
3654 CGGCCACCCGTCGCCACACACCTCTGCCACCCCATCTG..... 3696
902 erLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeuAlaGly 918
3697 .....ACGCGAAGAGAGACCCCTGTTCTCTGGCTGAT 3729
919 LysThrGlnThrAlaGlyValGluGluArgThrLysSerGlnLysProLe 935
3730 TCCGTATATCCCCAAGGCTTG.....CCTCCTCCTCT 3761
935 uProLysAspLysGluThrAsnArgAsnHisSerAsnSerAspAsnThrG 952
3762 CCGCAGCAGCAG..... 3774
952 LuThrLysAspProAlaProGlnAsnSerAsnProAlaGluProArg 968
3775 .....CCACAGCAGCGCCCTCAGCAGCCCGCCGAG 3807
969 LysProGlnLysSerGlyIleMetLysGlnArgValAsnProThrTh 985
3808 CAGCCCAAG.....TCCCTGATGACCAAGCTGCAAGGCGTAGTCACAA 3851
985 rAlaAsnSerAspLeu.....AsnProGlyT 994
3852 CTTGCGTTCGGGATTCACAGATTTCATGCGGTGCGGAGCCCGGGA 3901
994 hrThrGlnMetLysAspAsnPheAspIleGlyValCysProTrpGlu 1010
3902 CACCA.....GGAAC.....AGC 3915
1011 ValTyrAspLeuThrProGlyProValProSerGlu.....SerLysVa 1025
3916 CTGCGTCTCTGTACCGCCCGCCCTCCCGCAACACCTGCAGATGCT 3965
1025 lGlnLysHisValSerIleValAlaSerGluMetGluLysAsnProThrP 1042
3966 GCCCTGCACCTGAGCACCTTCCAGGAGGAGTCCATCTCCCTCCCTGGGG 4015
1042 heSerLeuLysGluLysSerHisHisLysProLysAlaAlaGluValCys 1058
4016 AGCACATCGATGATGACAGT..... 4035
1059 GlnGlnSerAsnGlnLysArgIleAspLysAlaGluValCysLeuTrpG 1075

```


1452 TGGACACACACAGGAAGTAATCTTGGTTCCCTGAGTTCTGGCAACATCGC 1501
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu... 296
1502 TTCAGTGTCCCTACCTGGACACCTCTTGGAAAAACCCAACTTTAAGAA 1551
297ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
1552 AGTGTGCAGGAGGAATGAAGCTTGGAGAAAGTAATGTCCAGGACAGCA 1601
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
1602 AAATGGATTGTTCATCAATGCCATCTATGCCATGGCACATGGGCTCCAG 1651
328 sLysCys..His.....LeuAsnAsnSerGluCysMe 338
1652 AACATGCACCATGCTCTGTGTCCGGGCCATGTGGGCTGTGTGATGCTAT 1701
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
1702 GAAACCATTTGATGGCAGGAAGCTCCTGGATTTCCTCATCAATTCCTCTT 1751
347 347
1752 TTGTCCGAGTCTCTGGAGAGAGGTGTGGTTTCATGATGAGAGGGGATGCT 1801
348 ...GlyAlaTyrGluCysIle..... 353
1802 CCCGAAGGATGACATTATCAATCTCAGTACACAGAAGCTAATCGCTA 1851
354CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
1852 TGACTATGTCCACGTGGGACCTGGCATCTGAAGAGAGTGCTGAATATTGATG 1901
368 snPheArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
1902 ATTACAAAATCCAG.....ATGAACAAAGCGGAATGGTACGATCT 1942
385 ValserGluGlu..... 388
1943 GTGTGCAGTCAGCTCTTAAAGGGTCAGATTAAAGTTCATACGGAAGG 1992
389AlaTyrValCysLeuProCysArgGluGly..... 398
1993 AGAAGTGAGCTGTGTGTGGATCTGCACGGCTGCAAAAGAGAAATGAGTTG 2042
399CysProPheCys..... 402
2043 TGCAGGACGAGTTCACCTGCAGACCTGTGACCTGGGGTGGTGGCCCAAC 2092
403 AlaAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2093 GCAGAGCTCACAGCTGTGAGCCCATTCCTGCTCCGTTATCTTGAGTGGAG 2142
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
2143 TGACATAGAATCATCATAGCCATCGCTTTCTTGTGCTGGGCATCCTCG 2192
430 euAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSer 446
2193 TGACGCTGTGTTCACCTCATCTTCGTTCTGTACCGGGACACACCCGTG 2242
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLe 463
2243 GTCAATCTCCAGTAGGAGCTCTGCTATATCATTTCTGCTGCTGATATTT 2292
463 uLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
2293 CCTCGGCTATGTGCTGCTTTCACCTCATCGCCAAACCTTACTACCAT 2342
480 rgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
1452 CCTGCTACCTCCAGCGCTCTCTAGTTGGCTCTCTTCTGGCATGTGTAT 2392
497 GlyThrValThrLeuLysLeuHisArgValLeuLysVal..... 509
2393 TCTGCTTTAGTGACCAAAACCAATCGTATTGCCAGCATCTGGTGGCAG 2442
510PheLeuSerArgThrAlaG 516
2443 CAAGAAGAGATCTGCACCGGAAGCCAGATTCATGAGCGCTTGGGCC 2492
516 lnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
2493 AAGTATCATAGCCTCCATCTGATTAGTGTACAGCTAAACACTAGTGGTG 2542
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCy 549
2543 ACCTTGATCATC..... 2554
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerA 566
2555ATGGAGCCTCCCATGCCCATTTTGTCTACCCGAGTATCAAGG 2597
566 sphHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
2598 AA.....GTCTACCTTATCTGCAATACCAGCAACCTGGGTGTAGTGGCC 2641
583 AlaValAla...GluPheLeuLeuLeuTrpGlyValTyrLeuCysTy 598
2642 CCTGTGGTTACATGGACTCCTCATCATGAGCTGTACTACTATGCTT 2691
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlav 615
2692 CAAGACCCGCAACGTGCGGCCAACTTCAATGAGCTAAATACATCGCT 2741
615 alAlaValHisAsnGluLeuIleSerAlaIlePheHisThrIleArg 631
2742 TCACCTGTACACTACCTCATCTGCTGGTGGCTTTCGTTCCCATTTAC 2791
632 PheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPh 648
2792 TTT.....GGGAGCAACTACAAGATCATCATCTACCTG 2823
648 eAlaHisThrHisLeuThrValThrValThrIleGlyLeuLeuIleP 665
2824 CTTGCGGTGAGCTCAGTGTGCGTGGCCCTGGGTGCTGCTTTACTC 2873
665 rOLys.....PheSerHisSerSerAsnAsnProArgAspAsp 677
2874 CGAAGATGTACATCATCTTCCCAACCTGAGAGGAACGTCGCGAGTGCC 2923
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
2924 TTCAGGACC.....TCTGATGTTGTCGCGATGCAGCTGCGGTGATGG 2964
691ArgSerGlySerTyrLeuAsnSerSerIleAsnSerA 703
2965 CAAACTGCCGTGCGCTCCCAACACTCTCTCAAC..... 2998
703 laTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLys 719
2998 2998
720 LysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAs 736
2999ATTTCCGGAGAAAGAGCCCGGGGAGG 3027
736 nAsnProHisLeuGlnLysLysArgCysSer..... 746
3028 GAATGCCAATTCATAACGGCAAGTGTGTGTCATGCTGTGAACACAGTGGAA 3077
747 ..LysLysGlyLeuGlyArgSerIleMetArgIleThrGluIlePro 762
3078 GACAGCGGCCCAAGGACAGCAGCTGTGGAGCGCTCTCTGTGCACGTG 3127

```

763 GluThrValSerArgGlnCysSerLysGluAspLysGluGluValAlaAspHis 779
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3128 AAGACCAACGAGCGCCGTAAACAA..... 3154
779 sglyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerS 796
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3155 .....ACAGCGGTAAATCAAA..... 3169
796 erGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPhe 812
3169 ..... 3169
813 SerLeuLysLysSerHisSerThrThrAspHisValArgAspGlnThrGln 829
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3170 CCCCTCACTAAAGTTAC..... 3187
829 uGluSerSerLeuProThrGluSerGlnGluGluGluThrThrGluA 846
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3188 .....CAAGGCTCTGGCA 3200
846 snSerThrLeuGluSerLeuSerGlyLysLysLeuThrClnLysLeuLys 862
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3201 ACAGCCTGACCTTTTCAGATGCCAGCACCAAGACCCCTTTTACAAATGTGGAA 3250
863 GluAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSe 879
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3251 GAAGAGGACAATACCCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 3300
879 rAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgT 896
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3301 TATGTTGTTGTCACCGAGCGGCCACCGCTGGCCACACACACCTCTGC 3350
896 hrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysLys 912
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3351 CACCCCATCTG.....ACCGCAAGAGAGACC 3376
913 ThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgTh 929
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3377 CCCCTGTTCTGGCTGATTCGCTCATCCCAAGGCGTTG..... 3415
929 rLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisS 946
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3416 .....CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3439
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsn 962
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3440 .....CCACAGCAGCGCGCC 3454
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGln 979
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3455 CCTCAGCAGCCCCCGCAGCAGCCCAAG.....TCCCTGATGGACCACT 3498
979 nArgValAsnProThrThrAlaAsnSerAspLeu..... 990
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3499 GCAAGGCGTAGTCACCAACTTCGGTTCGGGATTCAGATTTCATGCCG 3548
991 .....AsnProGlyThrGlnMetLysAspAsnPhe 1001
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3549 TCCTGGCAGCGCGCGGACACAGGAAACAG..... 3580
1002 AspIleGlyGluValCysProThrGluValTyAspLeuThrProGlyPr 1018
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3581 .....CTGGCCTCTCTGTACCCCGCCCC 3603
1018 oValProSerGlu.....SerLysValGlnLysHisValSerIleValA 1033
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3604 GCCTCCGCGCAACACCTCGAGATGCTGCCCTGCACCTGACGACCTTCC 3653
1033 laSerGluMetGluLysAsnProThrPheSerLeuLysLysSerHis 1049
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3654 AGGAGGATCCATCTCCCTCTCTGGGAGGACATCGATGATGACAGT... 3700

```

```

1050 HisLysProLysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgI 1066
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3701 .....GAGAGATT 3708
1066 easpLysAlaGluValCysLeuTrpGluSerGlnGly.....GlnS 1080
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3709 CAAGCTCTCTGTCAGAGATTTCGTGTACGAGCGCGAAGGACACCGAAGAG 3758
1080 erIleLeuGluAspGluLysLeuIleSerLysThrProValLeuPro 1096
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3759 ATCAATTGGAAGAGGAGGAGGACCTGCCACAGCAGCAGCTGACCCCT 3808
1097 GluArgAlaLysGluGluAsnGlyGlnProArgAlaAlaAsnValCy 1113
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3809 GAGGATTCTCTCTGCTGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3858
1113 sLaGlyGlnSerGluGluLeuProPro.....LysAlaValAlaSerL 1128
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3859 CTCTGGCAGCTCAGTCCGCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCT 3908
1128 ysThrGluAsnGluAsnLeuAsnGlnIleGlyHisClnGlnLysLysThr 1144
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3909 CCCCTCCAAATGTAACCTACGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3958
1145 SerSerSer 1147
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
3959 AGCTCTTCC 3967

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31058

seq_documentation_block:

ID_AAZ31058 standard; cDNA; 3831 BP.

XX AC AAZ31058;

XX DT 07-JAN-2000 (first entry)

XX DE phCar/hmglur2*Gq15 fusion construct cDNA sequence.

XX KW G-protein fusion receptor; CaR; calcium receptor; GluR; head injury;
 KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
 KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
 KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
 KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
 KW cognitive disorder; ss.

XX OS Homo sapiens.

XX PN WO951641-A1.

XX PD 14-OCT-1999.

XX PF 02-APR-1999; 99WO-US07333.

XX PR 03-APR-1998; 98US-0080671.

XX PA (NPSP-) NPS PHARM INC.

XX PI Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;
 PI Simin RT;

XX DR WPI; 1999-610995/52.

XX DR P-PSDB; AAY49127.

XX PT New G-protein fusion receptors and chimeras containing domains from
 PT different receptors, used to screen for modulators, potentially useful
 PT e.g. for treating or preventing stroke or Alzheimer's disease

XX PS Example 1; Fig 11; 255pp; English.

XX CC The invention relates to G-protein fusion receptors (I) comprising:

XX CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
 XX CC intracellular (ICD) domains, each chosen independently from a CaR

CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
 CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
 CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
 CC linker. (1), and recombinant chimeric receptors (CR) without the GP
 CC component, are used to assess function of the various domains and to
 CC identify compounds (e.g. allosteric modulators or antagonists) that act
 CC on these domains. The modulators are potentially useful for treating or
 CC preventing diseases associated with the receptors, e.g. stroke, head or
 CC spinal cord injury, epilepsy, ischemia, hypoglycemia, anoxia,
 CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
 CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
 CC (1) for recombinant production of corresponding proteins; and (2) to
 CC produce cells used in screening for modulators. Use of Car and mGluR
 CC domains allows presentation of GABABR domains, to a binding agent, in a
 CC form more like the natural domain structure compared with use of
 CC incomplete receptors, lacking one or more domains. By shuffling different
 CC domains, agents can be identified that affect particular domains of a
 CC receptor.

xx
 SQ Sequence 3831 BP; 917 A; 1025 C; 998 G; 891 T; 0 other;

alignment_scores:

Quality: 217.50 Length: 869
 Ratio: 0.583 Gaps: 38
 Percent Similarity: 42.923 Percent Identity: 20.253

alignment_block:

US-09-775-181-2 x AA231058 ..

Align seg 1/1 to: AA231058 from: 1 to: 3831

```

75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspse 91
   :::::|||||
394 TCAGAGCACATTCCTCTAGGATTGCTGTGGTGGGAGCAACTGGCTCAGG 443
   |||
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   |||||
444 CGTCTCCACGGCAGTGGCAAAATCTGTGGGGCTCTTACATTT..... 486
   |||
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   |||
487 .....CCCCAGTCAGTTATGCTCCCTCCACG.....AGA 516
   |||
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
   |||
517 CTCCTCAGCAACAAGATCAATCAAGTCTTCTCCGGAACC...ATCCC 563
   |||
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
   |||
564 CAATGATGACACACCGCCACTGCCATGGCAGACATCATCGATATTTC 613
   |||
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   |||
614 GCTGGAACCTGGTGGGCACAAATTGCAGCTCATGACGACTATGGGGCGCG 663
   |||
173 .....ArgAlaAlaIleThrPheSerT 180
   |||
664 GGGATTGAGAAATTCGAGAGGAGTGGAGAAAGGGATATCTGCATCGA 713
   |||
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   |||
714 CTTCACTGAACATCTCCCACTCTGATGAGAGAGATCCAGCATG 763
   |||
196 .....ArgGluGluSerArgIleLeuLeuG 204
   |||
764 TGGTAGAGGTGATTCAAAATTCACGCCCAAGTCACTCGTGGTTTCTCC 813
   |||
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuLeuThr 220
   |||
814 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATTGTCGGCGCA... 860
   |||
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237

```

```

861 .....TATCAGGGGCAAGATCTGGCTGGCCAGGAGGCTGGGCCAGCT 904
   |||
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   |||
905 CCTCCCTGATCGCATGCTCAGTACTTCCACGTGGTGGCGGCACCAT 954
   |||
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   |||
955 GGATTCGCTCTGAAGG.....TGGGCA 977
   |||
267 ProProTyrLeuGlu.....Cy 272
   |||
978 GATCCAGGCTTCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1027
   |||
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   |||
1028 TCCACAATGCTTTTCCCAAGGAGTTTGGGAAGAACAACATTTAACTGCCAC 1077
   |||
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   |||
1078 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTGAGAGG 1127
   |||
295 .....AsnLeuValProGluPheArgG 302
   |||
1128 TCAGGAAGAAGTGGCGACAGGTTTACCAACAGCTCGACAGCCTTCCGAC 1177
   |||
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   |||
1178 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGCCCTTACATA 1227
   |||
313 ..... 313
   |||
1228 GATTACACGCAATTTACGGATATCCTACATGTGTACTTAGCAGTCTACTC 1277
   |||
314 .....AspIleAspGlnCys..... 318
   |||
1278 CATTCGCCACGCTTGCAGATATATATATACCTGTACCTGGGAGAGGGC 1327
   |||
319 .....SerSerAspGlyTrpPhe 324
   |||
1328 TCITCACCANTGCTCTGTGCGACACATCAAGAAAGTTGAGCGCTGGCAG 1377
   |||
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   |||
1378 GTCTGAAGCACCTACGCGCATCTAAACTTTACAAACAATATGGGGAGCA 1427
   |||
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
   |||
1428 GGTGACCTTTGATGAGTGTGTGACCTGTGGTGGGAACTATTCCATCATCA 1477
   |||
348 .....Gly 348
   |||
1478 ACTGGCACCTCTCCCGACAGAGATCGCTCCATCGTGTTTAAGGAAGTCGG 1527
   |||
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG 362
   |||
1528 TATTACAACGTCTATGCCAAGAAGGAGAAAGACTCTTTCATCAACAGAGA 1577
   |||
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   |||
1578 GAAATCTCTGGAGTGGTGTCTCCAGGAGGAGTGCCTCTTCCAACATGCA 1627
   |||
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   |||
1628 GCCCAGACTCCTGGCAGGACCAAGGAAAGGATCATTTAGGGGGAGCC 1677
   |||
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   |||
1678 ACCTGCTGCTTGTAGTGTGTGGAGTGTCTGTATGGGAGTATAGTGATGA 1727
   |||
399 .....CysProPheCysAlaAspASP..... 405
   |||

```

1728 GACAGATGCCAGTGGCTGTAAACAAGTGGCCAGATGACTTCTGGTCCAATG 1777
406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
::: ||||| ::::: |||||
1778 AGAACACACACCTCTGCTTGAAGATGCCAGGAGTACATCGCTGGGCG 1827
418LeuAlaIleLeuSerPheGlnGlyLeuCy 427
::: ||||| ::::: |||||
1828 GATGCTGGGCTGTGGGAGCTGTACCATCGCTGCCTGGTGGCTGGC 1877
427 smetLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysA 444
1878 CACCTG.....TTTGTCTGGGTGCTTTGTGGGCAC...AATGCCA 1918
444 laLysSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPhe 460
::: ||||| |||||
1919 CACCAGTGTCAAGCCTCAGGTGGGAGCTGTGTACATCGCTGGGT 1968
461 GlySerLeuLeuTyrPheProValValIleLeuTyrPheGluProSe 477
||| ::::: ||| ::::: |||||
1969 GGTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2018
477 rThrPheArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAla 494
||| ||||| |||||
2019 CACGCGAGTGTACCTTACGCGGTGCTGGTGGGACCTGCTGCTCTCTG 2068
494 hrValTyrGlyThrValThrLeuLysLeuHisArgValLeuLysValPhe 510
::: ||||| |||||
2069 TCTGCTACTCAGCCCTGCTCACCACCAACACCGCATTCGACGATCTTC 2118
511LeuSerArgThrAlaGlnArgIleProTyrMetThr...GlyCl 524
::: ||||| |||||
2119 GGTGGGCGCGGAGGTGCCAGCGGCAGCGCTTCATCAGTCTGCTGCTC 2168
524 vArgValMetArgMetLeuAlaValIle.....LeuLeuValV 537
2169 ACAGTGCCCATCTCCCTGGCATTATCTCGGGCAGCTGCTCAATCGTGG 2218
537 alPheTrpPheLeuIle.....GlyTrpThrSerSerValCysGln 550
|| ||||| |||||
2219 TCGCCTGGCTGGTGGGAGCAGCGGCACAGGCAAGGAGACAGACGCC 2268
551 AsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAsp... 566
::: ||||| ::::: |||||
2269 GAACGGCGGAGGTGGTGCATCTGCTGCAACACCGCATGCAAGTAT 2318
567HisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyr 581
||| ||||| |||||
2319 GTTGGGCTCGTGGCTACAATGTCTCTCATC..... 2352
581 etThrAlaValAlaGluPheLeuPheLeuTrpGlyValTyrLeuCys 597
2353GGCCTCGCAGCTTTAT...GCC 2373
598 TyrAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAl 614
::: ||||| |||||
2374 TCAATACCTCGCAGTGGCCGAAACCTTCAACGAGGCCAAGTTCATTGG 2423
614 aValAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleA 631
::: ||||| |||||
2424 CTTACCATGTACACCATCTGCATCATCTGGCTGGCATTTGTTGCCATCT 2473
631 rgPheValLeuAlaSer.....ArgLeuGlnSerAspTrpMetLeuMet 645
::: ||||| |||||
2474 TCTATGTACCTCCAGTACTACCGGGGTACAGACCAACCATCGTGGCTG 2523
646 LeuTyrPheAlaHisThrHisLeuValThrValThrIleGlyLeuLe 662
::: ||||| |||||
2524TCAGTCAGCTCAGCGGCTCCGTGGTGGTGGCTGCCT 2561
662 uLeuIleProLysPhe..... 667
2562 CTTTGGCCCAAGCTGCATCATCTCTTCCAGCGCGGAGAACGATGG 2611

668 ..SerHisSerAsnAsnProArgAspIleAlaThrGluAlaTyr 683
||||| ::::: |||||
2612 TTAGCACCGGGCACCACCGAGCTTTGGCAGTGTCTGCCAGGCGC 2661
684 GluAspGluLeuAspMetGlyArgSerGlySerTyr..... 695
::: ||||| ::::: |||||
2662 AGCTCAGCCTTGGCAAAGGTCTGGCTCCAGTTGTCTCCACCTGTGTG 2711
696LeuA 697
2712 CAATGGCGGTGAGTGGTGGACTCGACAACGTCATCGCTTATGACTCTGG 2761
697 snSerIleAsnSerAlaTrpSerGluHisSerLeuAspProGluAsp 713
::: ||||| ::::: |||||
2762 AGTCATCATGCGTGTGCTGAGCGAGGAGGCCAAGAACCGCGCGG 2811
714 IleArgAspGluLeuLysLysLeuTyrAlaGlnLeuGluIleTyrLysAr 730
||||| ::::: |||||
2812 ATCAACGACGAGATCGAGCG.....CAGCTCCGAGGAGCAAGCG 2852
730 g 730
2853 G 2853
seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ10801
seq_documentation_block:
ID AAZ10801 standard; DNA; 3591 BP.
XX AC AAZ10801;
XX DT 09-NOV-1999 (first entry)
XX DE Mouse metabotropic glutamate receptor 5 (mGluR5) gene.
XX KW Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;
KW depressant; anaesthesia; phospholipase C; intracellular calcium; ss.
XX OS Mus sp.
XX FH Key Location/Qualifiers
FT CDS 1..3591
FT FT /*tag= a
FT FT /product= mGluR5
FT FT /note= "metabotropic glutamate receptor 5"
XX PN W09938975-A2.
XX PD 05-AUG-1999.
XX PE 29-JAN-1999; 99WO-US02033.
XX PR 30-JAN-1998; 98US-0073190.
XX PA (UYTE-) UNIV TECHNOLOGY CORP.
XX PI Johnson TE, Rikke BA, Sikela JM, Simpson VJ;
XX WPI; 1999-494093/41.
XX DR P-PSDB; AAY28562.
XX CNS depressant sensitivity associated a YAC containing murine
PT tyrosinase and metabotropic glutamate receptor 5 genes, useful for
PT identifying CNS depressant modulators
XX Claim 21; Fig 14; 140pp; English.
XX This sequence is the mouse metabotropic glutamate receptor 5 gene
CC (mGluR5). mGluR5 are G protein-coupled receptors capable of activating a
CC variety of second messenger systems. mGluR5 belongs to mGluR Group I. The
CC binding of an agonist to group I receptors results in the activation of
CC phospholipase C causing the mobilisation of intracellular calcium. The

CC mouse mGluR5 sequence correspond to parts of the mouse YRT2
 CC polynucleotide which is associated with central nervous system (CNS)
 CC depressant activity. CNS depressants are used for a variety of purposes
 CC including anaesthesia.
 XX
 SQ Sequence 3591 BP; 911 A; 913 C; 939 G; 828 T; 0 other;

alignment_scores:
 Quality: 214.00 Length: 964
 Ratio: 0.495 Gaps: 37
 Percent Similarity: 44.813 Percent Identity: 18.465

alignment_block:

US-09-775-181-2 x AA210801 ..

Align seg 1/1 to: AA210801 from: 1 to: 3591

```

219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuH1 235
   :::::::::::::::::::: ||| ||| ||| ||| ::::
994 GATCTCAAGTGGTTGATGATTAATCTGAAGCTCCGGCCAGAAACAAA 1043
   sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg. 251
   : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
1044 CCTCAGAAACCTT ..... TGGTTTCAAGAATTTTGGCAGC 1078

252 ..... ArgLysAspGlyLeuGlyGlyAspLysSerHisPhe 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1079 ATCGTTTTTCACTGCGCGTAGAAGGTTTGCACAGACGAACGAAGTAC 1128

264 LysTrpSerProTrpLeuGluCysGluAsnGly ..... 275
   ::::: |||: |||: |||: |||: |||: |||: |||: |||: |||:
1129 AACAAAGACT ..... TGAACAGATTCTCTAACTCTGAGAAC 1163

276 ..... SerTrpLysProGlyTrpLeuVal... ThrLeuSerS 287
   :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1164 GCATCATGTTCAAGATTCCAAAATGGGATTTGTGATCAATGCAATCTATT 1213

287 erAlaIleTyrrGlyLeu ..... GlnProAsnLeuValProGluPhe 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1214 CTATGGCTTATGGGCTCCACAAACATGCAGATGCCCTGTGTCAGGCTAT 1263

301 ArgGlyVal ..... MetLysValAspI1 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1264 GCAGGCTCTGTGATGAATGAAGCCAAATGTATGGCGGAACATTTTGA 1313

308 eAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPheS 325
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1314 CTCCTCATGTAACCAACTTTACTGGAGTTCCGGAGATATGATTTCTAT 1363

325 erGlyThrHisLysCysHisLeuAsnSerGluCysMetProIleLys 341
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1364 TTGATGAAATGGAGACTCTCCAGGAAGGTATGAATAATGAATTTCAAG 1413

342 GlyLeuGlyPheValLeuGlyAlaTyrrGluCysIleCysLysAlaGlyPh 358
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1414 GAAATGGGA ..... AAAGATATTATTGATTACATCAATGTTGAAG 1454

358 eTyrrHisProGlyValLeuProVal ..... 366
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1455 TTGGCAATGGGGAATTAATAATGGATGATGATGAAGTGGTCCAGA 1504

367 ..AsnAsnPheArgArgGlyProAspGlnHisIle... SerGlySer 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1505 AAAATAACATCATCATCTCTGTGTCAGTGAACCGTGTGAGAAAGGACAG 1554

382 ThrLysAspValSerGluGlu ..... AlaTyrrValCysLeuPr 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1555 ATAAAGGTGATCCGGAAGGAGGAGTACAGCTGTTGTTGGACCTGCACACC 1604

394 oCysArgGluGly ..... CysProPheCys. 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1605 TTGTAAGAGATGAGTATGTTTGTGATGAGTACACCTGCANAGCGGTGCC 1654

```

```

403 ..... AlaAspAsp... SerProCysPheValGlnGlu 412
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1655 AACTGGGGTCTGCGCCACTGACGACTTACAGGTGTGATTGATTCACA 1704

413 AspLysTyrrLeuArg ..... LeuAlaIleIleSerPh 423
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1705 GTCCAGTATCTTCGATGGGGTGACCCCTGAGCCCATTCAGCTGGTGT 1754

423 eGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValValTyrrH 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1755 TGCTGCTCGTCTGCTAGCCACCCTCTGCTTACTGTAATCTTCATCA 1804

440 isPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuGlu 456
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1805 TTTATCGGACACTCCAGTGGTCAAGTCTTCAGCAGGGAACCTCTGCTAC 1854

457 ThrIleLeuPheGlySerLeuLeuLeuTyrrPheProValValIleLeuTy 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1855 ATTATCTTGTCTGGCATCTGCTGGGTACTTGTGTACCTTCTGCCCTCAT 1904

473 rPheGluProSerThrPheArgCysIleLeuLeuLeuArgTrpAlaArgLeuL 490
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1905 TGCAGAGCCCAACAGATTTACTGCTACCTTCAGAGATTTGGCATCGGTC 1954

490 euGlyPheAlaThrValTyrrGlyThrValThrLeuLysLeuHisArgVal 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1955 TCTCTCCAGCCATGAGCTACTCAGCCCTTGTACGAAAGACCAACCGTATT 2004

507 LeuLysVal ..... 509
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2005 GCAAGGATTTCTAGTGGCAGCAAGAAGAATCTGTACCAAAAAGCCAG 2054

510 .PheLeuSerArgThrAlaGlnArgIleProTyrrMetThrGlyGlyArgV 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2055 ATTCATGAGCCCTGTGCTCAGTTAGTG ..... A 2083

526 alMetArgMetLeuAlaValIleLeuLeuValValPheTrpPheLeuI1 542
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2084 TCGCTTTCATTTCTCATCTGTATACAGTTGGGTATTATTGTGGCCCTCTTT 2133

542 eGlyTrpThrSerSerVal... CysGlnAsnLeuGluLysGln ..... 555
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2134 ATCATGAGCCCTCCGGATATAATGATGACTATCCAGCATCCGAGAGAT 2183

556 ..... IleSerLeuIleGlyGln 561
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2184 CTACTTGTATTGTAAACACCACCACTAGGGGTGTCTACTC ..... 2224

562 GlyLysThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTr 578
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2225 ..... CTCTTGGATACAAT 2238

578 pAspTyrrMetThrAlaValAlaGluPheLeuLeuLeuTrpGlyValT 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2239 GGATTATTG ..... ATTTGAGTTGCACAT 2263

595 yrLeuCysTyrrAlaValArgThrValProSerAlaPheHisGluProArg 611
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2264 TCTATGGGTTCAAGCCAGAAATGTTCCAGCCAACTTTAAGAGGCCAAA 2313

612 TyrMetAlaValAlaHisAsnGluLeuIleIleSerAlaIlePheHi 628
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2314 TATATTGCTTTCAACATGTACACCAACCTGCATATAATATGCTGGCTTTGT 2363

628 sThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuM 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2364 GCCTATCTACTTT ..... GGCAGCAACTACAAAATCA 2395

645 etLeuTyrrPheAlaHisThrHisLeuThrValThrValIleGlyLeu 661
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2396 TCACCATGTGTTTCTCAGTCCAGCTCAGTCCACAGTGGCCCTGGGTGC 2445

```

```

662 LeuLeuIleProLys.....PheSerHisSerSerAsnAsnPr 674
|||||
2446 ATGTTTGTCCGAAGGTGTACATCATCTAGCCAGCAACCGGAGAGAAATGT 2495
674 oArgAspAspIleAlaThrGlu...AlaTyrGluAspGluLeuAspMetG 690
|||||
2496 GCGCAGCGGCTTCACAACTCTACAGTGTGTGCGCATCGCAGTAGGAGATG 2545
690 lYargSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlaTrp 704
|||||
2546 GCAAGTCATCATCCGCTGCCAGCAGATCCAGCAGCCTAGTCAACCTGTGG 2595
705 SerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLysLe 721
|||||
2596 AAGACGAGGGGCTCGTCTGGGAAACCTTAAGGTACAAACAGCAGAGACT 2645
721 u.....TyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleT 735
|||||
2646 GCGCCAGCACAGTCGGAATAGAGTGTTCACCCCAANGGAGGTATGG 2695
735 hrAsnAsnProHisLeuGlnLysLysArgCysSerLysLysGlyLeu... 750
|||||
2696 GGAATGGTGGAGAGCAACAATGAGCAGCTCCCAATGGAATAATCCGTGACT 2745
751 .....GlyArgSerIleMetArg... 757
|||||
2746 TGGGCCAGCAATGAGAAGAGACACCGGGGGCAGCACCTGTGGCAGCGACT 2795
758 .....IleT 759
2796 GTCGTCCATATCACAGAAGAGAGACCCCAACCCAGCAGCAGTCACTCA 2845
759 hrGluIleProGluThrValSerArgGlnCysSerLysGluAspLysGlu 775
|||||
2846 AACCTCTCCCAAGAGCAGAG.....AGCCGGGGCAGGGTGCA 2886
776 GlyAlaAspHisGlyThrAlaLysGly..... 784
2887 GGGGACAGTGGTGGCTCTGGCCCGCGTGAGTGGTGGTAGCGCAGG 2936
785 ....ThrAlaLeuIleArgLysAsnProProGluSerSerGlyAsnThrG 800
|||||
2937 ATGCACAGGACAGGGGGCCAGACCCAGCAGCCGCGCCCAAGGGC. 2985
800 lYlsSerLysGluGluThrLeuLysAsnArgValPheSerLeuLysLys 816
2985 ..... 2985
817 SerHisSerThrTyrAspHisValArgAspGlnThrGluGluSerSerSe 833
|||||
2986 .....CTTTATGATGTCGACAGAGCGCAGAGCGC 3014
833 rLeuProThrGluSerGlnGluGluThrThrGluAsnSerThrLeuG 850
|||||
3015 CTTCCAGCGGCTGCCAGCGCGCTGCGCCATCGCCCATCAGTACGCTGA 3064
850 luSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAspSerGlu 866
|||||
3065 GCCACCTGGCAGGC.....TCGGCG 3084
867 AlaGluSerThrGluSerValProLeuValCysLysSerAlaSerAlaHl 883
|||||
3085 GCGCCACAGCAGCAGCAGCGCGCTGCGTGCACCTCGGACCGCTGCAGC 3134
883 sAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMetLeuG 900
|||||
3135 CAGCAGCTCATCCCGGGCTCGCTCATGAGCAGATTAGCAGCGTGGTGA 3184
900 lNlysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeu 916
|||||
3185 CGCGCTTACCGCCCAACATCACCAGCACTCACTCCATGATGTGTGCCACC 3234
917 AlaGly..... 918

```

```

|||||
3235 GCGGCTGCGCGCGGCCCCCTGGTACCCCTATCTGCTCTTCTTCTACTGAT 3284
919 .....LysThrGlnThrAlaGlyValGluGlu 928
|||||
3285 CCCCRAAGAGATCCAGTCCGCCACGACCATGACGCTTCGCAGAG.... 3330
928 rgThrLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsn 944
|||||
3331 .....ATCCAGCCTCTGCGGCCCATCGAGTGACC..... 3360
945 HisSerAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSe 961
|||||
3361 .....GGAGGAGCTCAGCCGCGCAGCAGGCGCATCATCTGCCAAGA 3401
961 rAsnProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysG 978
|||||
3402 GACGCCGCGCAGGAGCTGAAGCGCCCGCCAGGAACCGGATCTGGAGGAGC 3451
978 lNlArgValAsnProThrThrAlaAsnSerAspLeuAsnProGlyThr 994
|||||
3452 TGGTGGCCCTCACTCCA.....CCA 3471
995 ThrGlnMetLysAspAsnPheAspIleGlyGluValCysProTrpGluVa 1011
|||||
3472 TCGCCCTTCAGGAGCTCGGTGGACTCGGGGAGCAGCACCCCAACTCTCC 3521
1011 lTyrAspLeuThrProGlyProValProSerGluSerLys 1024
3522 AGTCCTCCGAATCGGCCCTCTGCATCCCATCTCTCCCAAA 3561

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:AAQ80419
seq_documentation_block:
ID: AAQ80419 standard; cdna; 4085 BP.
XX
AC AAQ80419;
XX
DT 21-JUL-1995 (first entry)
XX
DE Human mGluR5a cDNA.
XX
KW Metabotropic glutamate receptor; mGluR5; mGluR5a; hippocampus;
cerebellum; pCMV-T7-3; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 370..3912
FT FT /*tag= a
FT FT /product= human mGluR5a
XX
PN WO9429449-A.
XX
PD 22-DEC-1994.
XX
PF 03-JUN-1994; 94WO-US06273.
XX
PR 04-JUN-1993; 93US-0072574.
XX
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
XX
PI Daggett L, Ellis SB, Hess SD, Johnson EC, Liaw C;
PI Pontsler A;
XX
DR WPI; 1995-036478/05.
DR P-PSDB; AAR64253.
XX
XX New DNA encoding human metabotropic glutamate receptor
sub: type(s) - and related proteins, probes, RNA, transformed
cells and antibodies, useful in treatment, diagnosis and
PT identification of specific modulators
XX

```



```
314 .....AspIleaspGlnCys..... 318
1278 CATTTGCCACGCTTCAAGATATATACCTGCTTACCTGGAGAGGGC 1327
319 .....SerSerAspGlyTpphe 324
1328 TCTTCCACCAATGCTCTCTGTGCAGACATCAAGAAATTGAGGCGTGGCAG 1377
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1378 GTCTCTCAAGCACCTACCGCATCTAAACTTTTACAACAATATGGGGAGCA 1427
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1428 GGTGACCTTTGATGAGTGTGGTACCTGTGGGGAACCTATTCCATCATCA 1477
348 .....Gly 348
1478 ACTGGCACCTCTCCCCAGAGGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1527
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
1528 TATTACAAGCTCTATGCCAAGAGGAGAAAGACTCTTTCATCAACAGGGA 1577
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
1578 GAAATCTCTGTGGAGTGGTGTCCAGGAGGTGCCCTTCTCCAACTGCA 1627
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
1628 GCCGAGACTGCTGGGAGGACCAGGAAGGATCATTTGAGGGGAGGCC 1677
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
1678 ACCTGCTGCTTTGAGTGTCTGAGTGTCTGATGGGAGTATAGTATGATGA 1727
399 .....CysProPheCysAlaAsp..... 405
1728 GACAGATGCCAGTGGCTGTAAAGTGGCCAGATGACTTCTGGTCCAATG 1777
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
1778 AGAACACACCTCTCTGACTGACAGTCCCGGAGGATACATCCGCTGGGGC 1827
418 .....LeuAlaIleIleSerPheGlnCysLeuCy 427
1828 GATGCTGGGCTGTGGGACCTGTCCACCATCGCTGCTCGGTCCCTGCGC 1877
427 sMetLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysA 444
1878 CACCCCTG.....TTGTGCTGGGTGTCTTTGTGGGCGAC...AATGCCA 1918
444 lAlYsSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPhe 460
1919 CACCAGTGTCAAGCCTCAGTGGGAGCTGTGTACATCTGCTGGGT 1968
461 GlySerLeuLeuTyrPheProValValIleLeuTyrPheGluProSe 477
1969 GGTGCTCTCTCTGCTACTGATGATGACCTTCATCTTCATTCGCCAAGCATC 2018
477 rThrPheArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaT 494
2019 CACGGCAGTGTCTACTCTACCGCGCTCTGGTTTGGGCACTGCTCTCTCTG 2068
494 hrValTyrGlyThrValThrLeuLysLeuHisArgValLeuLysValPhe 510
2069 TCTGTCTACTCAGCTGTCTCACCAGACCAACCGCATTCACGACATCTTC 2118
511 .....LeuSerArgThrAlaGlnArgIleProTyrMetThr...GlyG1 524
2119 GGTGGGCGCGGAGGGTGCACCGGCGGACGCTTCATCAGTCCGCTC 2168
524 yArgValMetArgMetLeuAlaValIle.....LeuLeuValV 537
```

```
2169 ACAGGTGGCCATCTGCTGGCACCTTATCTCGGCCAGCTGCTCATCGTGG 2218
537 alPheTrpPheLeuIle.....GlyTrpThrSerSerValCysGln 550
2219 TGCCTGGCTGTGTGGAGGACCGGCGACAGGCAAGGAGACAGCCGCC 2268
551 AsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAsp.. 566
2269 GAACGGCGGAGTGTGTGACACTGCGCTGCAACACCGCGCATGCAAGTAT 2318
567 .....HisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrM 581
2319 GTGGCTCGCTGGCTACAAATGTCTCTCATC..... 2352
581 etThrAlaValAlaGluPheLeuPheLeuLeuTrpGlyValTyrLeuCys 597
2353 .....GCCCTCTGCACGCTTTAT...GCC 2373
598 TyrAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAl 614
2374 TTCAATACTCGCAAGTGCCTCCGAAACTTCAACGAGGCCAAGTTCATTGG 2423
614 aValAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleA 631
2424 CTTCACCATGTACACCATCATCTGCTGGCTGGCATTTGTGCCCATCT 2473
631 rgPheValLeuAlaSer.....ArgLeuGlnSerAspTrpMetLeuMet 645
2474 TCTATGTACCTCCAGTGTACTACCGGTACAGACCACCATGTCGGTG 2523
646 LeuTyrPheAlaHisThrHisLeuThrValThrValIleGlyLeuLe 662
2524 .....TCAGTCAGCTCAGCGCTCGTGGTGGTGGTGGTGGT 2561
662 uLeuIleProLysPhe..... 667
2562 CTTTGGCGCCAAAGTGCACATCATCTCTTCCAGCGCGCAAGAAGCGTGG 2611
668 ..SerHisSerSerAsnAsnProArgAspAspIleAlaThrGluAlaTyr 683
2612 TTAGCCACCGGCGCACCCAGCCCTTTGGCAGTGTGCTGCCAGGGCC 2661
684 GluAspGluLeuAspMetGlyArgSerGlySerTyrLeuAsnSerSerI 700
2662 AGCTCCAGCTTGGCACAAGGCTGCTGCCCTCCAGTTTGTCCCACTGTTTG 2711
700 eAsn 701
2712 CAAT 2715
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ10803
seq_documentation_block:
ID AAZ10803 standard; DNA; 3918 BP.
XX
AC AAZ10803;
XX
XX AC
DT 09-NOV-1999 (first entry)
XX
DE Rat metabotropic glutamate receptor 5 (mGluR5) gene.
XX
XX Metabotropic glutamate receptor; mGluR5; central nervous system; CNS;
XX depressant; anaesthesia; phospholipase C; intracellular calcium; ss.
XX
XX Rattus sp.
XX
XX Key Location/Qualifiers
XX CDS 268..3783
XX FT /*tag= a
XX FT /product= mGluR5
XX FT /note= "Metabotropic glutamate receptor 5"
XX
```



```

559 .....IleGlyGlnGlyLysThrSerAspHisLeuI 569
2400 TATCATGGAGCCTCCGGATATATGACTACTCAAGCATCCGAGAAG 2449
569 IePheAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaValAla 585
2450 TCTACTTGATTGTATACACCACCACTAGGGTGTGTACTCTCTTGA 2499
586 ...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysTyrAlaValAr 601
2500 TACAATGGATTATTGATTGTTGAGTTCACGTTCTATGCGTTTAAAGACCAG 2549
601 gThrValProSerAlaPheHisGluProArgTyrMetAlaValAlaValIn 618
2550 AAATGTTCCAGCCAACTTAAATGAGCCAAATATATTGCTTTCACCATGT 2599
618 isAsnGluLeuIleLeuSerAlaIlePheHisThrIleArgPheValLeu 634
2600 ACACACCTGCATCATATGGCTGGCCTTCTGTCCTATCTACTTT..... 2643
635 AlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPheAlaHisTh 651
2644 .....GGCAGCAACTACAAATATCATCACCATGTTGTTTTCAGT 2681
651 rHisLeuThrValThrValThrIleGlyLeuLeuLeuIleProLys.... 666
2682 CAGCCTCAGTGCACAGTGGCCTGGGTTGTCATGTTGTCGGAAGGTGT 2731
667 .....PheSerHisSerSerAsnProArgAspAspIleAlaThr 680
2732 ACATCATCTAGCCAAACCGGAGAAATGTGGCAGCCCTTCACAACC 2781
681 Glu...AlaTyrGluAspGluLeuAspMetGlyArgSerGlySerTyrIle 696
2782 TCTACAGTGTGACGTCATGTCATGAGGAGCGCAATCATCTGTCGCTGC 2831
696 u.....AnSerSerIleAsnSerAlaTrpSerGluHisSerLeuAsp 711
2832 CAGCAGATCCAGCAGCTAGTCAACTGTGGAAGAGAGGGGCTCTCTCTG 2881
711 roGluAspIleArgAspGluLeuLysLys..... 720
2882 GGAACCTTAGCTCCACGAGAAATCTGTACTTGGGCCCCAGCAATGAG 2931
721 .....LeuTyrAlaGlnLeuGluIleTyrLysAr 730
2932 AAGAGTACCCGGGGCAACATTTGTGGCAGCGACTGCTGTCCACATCAA 2981
730 gLysLysMetIleThrAsnAsnProHisLeuGlnLys.....LysA 744
2982 CAAGAAGGAGAACCCCAACCAACGGCGGTTCATCAAAACCTTTTCCCAAGA 3031
744 rGlySerLysLysGlyLeuGlyArgSerIleMetArgArgIleThrGlu 760
3032 GCACAGAGAACCCGGGG..... 3048
761 IleProGluThrValSerArgGlnCysSerLysGluAspLysGluGlyAl 777
3049 ...CCTGGTGCAGCGCAGTGGTGGCTCGGGTCCCGGTGTAGCTGGTGC 3095
777 asPheLysGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProG 794
3096 T.....GTAACGAGGAGTGCACACACAGAGGTGGCCCGGAGCCACCG 3139
794 luserSerGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArg 810
3140 AGCCGGTCCCAAGGG..... 3156
811 ValPheSerLeuLysLysSerHisSerThrTyrAspHisValArgAspGl 827
3157 .....CTTTATGATGT 3167
827 nThrGluGluSerSerLeuProThrGluSerGlnGluGluThrT 844

```

```

3168 GCGGAGGAGGAGGAGGAGCTTCCCGCGGCTGCCAGCGCGCTCGCCAT 3217
844 hrGluAsnSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLys 860
3218 CGCCCATCAGCAGCTTGAGCCACCTCGCAGGC..... 3249
861 LeuLysGluAspSerGluAlaGluSerThrGluSerValProLeuValCy 877
3250 .....TCGCGGGCGGCACAGATGATGATGCGCGCTCGTGCA 3287
877 slySerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisP 894
3288 CTCTGAGACAGCTGCTCGCAGCAGCTGTCGCCAGGGCTCGCTCATGGAGC 3337
894 roArgThrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLys 910
3338 AGATCAGCAGCTAGTGACGCGCTTCACAGCCAACATCTCCGAGCTCAAC 3387
911 GluLysThrLeuGlyLeuAlaGly.....LysThrGlnT 922
3388 TCCATGATGTTGTCACCGCGGCCCGCGGCGGCGGCTGGTACTCCAAT 3437
919 .....LysThrGlnT 922
3438 CTGCTCCTCTACTGATCCCCAAAGAGATTGAGCTGCCAGCAGCATGA 3487
922 hrAlaGlyValGluGluArgThrLysSerGlnLysProLeuProLysAsp 938
3488 CAGCTTCGCAGAG.....ATCCAGCGCTGCGCGGCATC 3522
939 LysGluThrAsnArgAsnHisSerAsnSerAspAsnThrGluThrLysAs 955
3523 -GAGGTGACC.....GGAGAGCTCAGGGGGCGACAGGCGTATC 3560
955 proAlaProGlnAsnSerAsnProAlaGluGlu...ProArgLysProG 971
3561 ACCTGCCAGGAGCGCCACAGAGAGCTGAATCCGCCCGCGGCAACACAG 3610
971 lnLysSerGlyIleMetLysGlnArgValAsnProThrThrAlaAsn 987
3611 ATCTGGAGGAGCTTGGGCCCTCACTCCACCATCGCCC..... 3648
988 SerAspLeuAsnProGlyThrThrGlnMetLysAspAsnPheAspIleGl 1004
3649 .....TTCAGGAGCTCGTGGACTCGGG 3671
1004 yGluValCysProTrpGluValTyrAspLeuThrProGlyProValProS 1021
3672 GAGCACCACCCCAACTCTCCAGTCTCAGAAATCGGCCCTCTCGATCCCAT 3721
1021 erGluSerLys 1024
3722 CCTCTCCCAAA 3732
seq_name: /SIBS1/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80420
seq_documentation_block:
ID: AAQ80420 standard; cDNA; 4181 BP.
XX
XX AAQ80420;
XX
XX 21-JUL-1995 (first entry)
XX
XX Human mGluR5b cDNA.
XX
XX Metabotropic glutamate receptor; mGluR5; mGluR5b; hippocampus;
XX cerebellum; pCMV-T7-3; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 370..4008

```


782 AlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyAs 798
3304 GGTCCGGGCTCCGAGCAGCGCCGCGGCCAGCGCGGCCGAGTCCCGACAGCG 3353
798 nThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuL 815
3354 CGGCCCCAAAGCG 3366
815 yLysSerHisSerThrTyrAspHisValArgAspGlnThrGluGluSer 831
3367CTGTATGATGTCGCCGAGCGCTGAG 3390
832 SerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerTh 848
3391 GAGCACTTCCGGCGCCGCGCGGCCGCTCACCGTCGCCCATCAGCAC 3440
848 rLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAspS 865
3441 GCTGAGCCACCGCGCGGCTCGGCCACGCGCACGGAG 3478
865 erGluAlaGluSerThrGluSerValProLeuValCysLysSerAlaSer 881
3479 ACATGTGCGCTCGCTGCATCGGAGCTGTGGCGCGACAGCTCTCTCG 3528
882 AlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMe 898
3529CAGGGCTCCCT 3539
898 tLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuG 915
3540 CATGGAGCAGATCAGCAGAGTGTGTCACCCGCTC 3573
915 lyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSer 931
3574ACGCCAACATCAGCGG 3591
932 GlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSe 948
3592CTCAACTC 3599
948 rAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAla 964
3600 CATGATGCTCTCCACCGCGCCCGCCGCGCGGTCTGGCGCCGCTCT 3649
965 GluGluProArgLysProGlnLysSerGly 974
3650 GCTGCTCTACCTGATCCCCAAGAGATCCAGTGTCCACGACCATGAGCG 3699
975IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspL 990
3700 ACCTTTGCCGAATCCAGCTCTCCGGCCATCGAAGTCACGGGGCGGC 3749
990 euAsnProGlyThrThrGlnMetLysAsp 999
3750 TCAGCCCGCGCAGGGCGCAGCGGTGGGAGCGCGCGGAGAGGCC 3799
1000 AsnPheAspTleGlyGluValCysProTtpGluValTy 1012
3800 CCGCGCGGTCGCCAGGCTCGCGCCGCCAACCCAGACCTCGAGGAGCTG 3849
1012 rAspLeuThrProGlyProValProSerGluSerLysValGlnLysHisV 1029
3850 GTGGCTCTACCCCGCGCTCCCTTCAGAGACTCGGTGACTCGGGGAG 3899
1029 aISerIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys 1045
3900 CACAACCCCAACTCGCCAGTGTCCGAGTCGGCCCTCTGTATCCGCTGT 3949
1046GluLysSerHisLysLysProly 1053
3950 CTCCTCCAAATGACACTTATCATAGAGATTCACATCAGCAGCTCTCTCG 3999
1053 sAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaG 1070

```

i      : : : : :
4000 TCGTTGTGAA..... 4009
1070 luValCysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLys 1086
    ||| : : : : :
4010 .....TGTCCCTGGAAGCAGC..... 4027
1087 LeuLeuIleSerLysThrProValLeuProGluArgAlaLysGluGluAs 1103
    ||| : : : : :
4028 .....CGCCCTCGCGTCG..... 4042
1103 nGlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluL 1120
    ||| : : : : :
4043 .....GAGCGGA 4049
1120 euProProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGln 1136
    ||| : : : : :
4050 GCCCCCG.....TGTTCCACACACACAT..... 4075
1137 IleGlyHisGlnGluLysLysThrSerSerSerGluGluAsnValArgG1 1153
    ||| : : : : :
4076 ...GGCAAGCATAGTCGCTGGTTACGGCCCGAGGGGAAGATGCCAAGGG 4122
1153 ySer 1154
    ||| : : : : :
4123 CACC 4126

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV21511

seq_documentation_block:

ID AAV21511 standard; DNA; 6755 BP.

XX AC AAV21511;

XX DT 17-AUG-1998 (first entry)

XX DE Staphylococcal bacteriocin BacRI operon.

XX KW BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;
 KW Moraxella bovis; infectious bovine keratoconjunctivitis; cancer;
 KW therapy; ds.

XX OS Staphylococcus aureus strain UT0007 (ATCC 55800).

XX PN W09812319-A1.

XX PD 26-MAR-1998.

XX PF 18-SEP-1997; 97WO-US16758.

XX PR 17-SEP-1997; 97US-0931999.

XX PR 19-SEP-1996; 96US-0710561.

XX PA (UNIV) UNIV KANSAS STATE RES FOUND.

XX PI Crupper SS, Iandolo JJ;

XX DR WPI; 1998-230316/20.

XX PT Therapeutic proteinaceous substances from *Staphylococcus aureus* -
 PT useful to inhibit growth of wide range of prokaryotic or eukaryotic
 PT cells, e.g. *Moraxella bovis* causing infectious bovine
 PT keratoconjunctivitis

XX PS Claim 2; Page 19-23; 38pp; English.

XX CC This polynucleotide comprises the bacteriocin BacRI operon of
 CC *Staphylococcus aureus* UT0007. The sequence of the BacRI operon
 CC was determined by N-terminal sequencing of purified BacRI peptide
 CC (see AAW54171), with back-translation and plasmid analysis. The
 CC BacRI operon includes the BacRI gene (see AAV21510), a homologue of
 CC the *cylM* gene of the cytotoxin operon of *Enterococcus faecalis*,
 CC whose function is involved in the maturation of pre-cytolysin,

CC an ATP-transporter gene, *bio1* and *bio2* genes related to
 CC lactococcal biosynthesis and modification, and a gene involved in
 CC immunity function. BacRI peptide can be produced by construction
 CC of an expression vector containing an oligonucleotide or operon
 CC coding for BacRI, and use of the vector to transform host cells for
 CC BacRI expression. The entire BacRI operon has been cloned into
 CC plasmid PUB110, and *Bacillus subtilis* transformants secreted the
 CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is
 CC active against many Gram-positive and Gram-negative organisms such
 CC as *Bordetella bronchiseptica*, *Pasteurella multocida* and
 CC *Staphylococcus aureus*; *Moraxella bovis*, causing infectious bovine
 CC keratoconjunctivitis, is especially sensitive. BacRI can also be
 CC used as an anti-cancer agent.

XX SQ Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U; 0 other;

alignment_scores:

Quality: 204.50 Length: 461
 Ratio: 0.867 Gaps: 20
 Percent Similarity: 51.193 Percent Identity: 22.993

alignment_block:

US-09-775-181-2 x AAV21511

Align seg 1/1 to: AAV21511 from: 1 to: 6755

```

717 GluLeuLysLysLeuTyAlaGlnLeuGluIleTyLysArgLysLysMe 733
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3609 GAAAAAATAAAACCCCAAGCACACACCGACGACCAAAAAA..... 3653
733 tIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysLysGlyL 750
    ||| : : : : :
3654 ....ACGCCGAAACCAACAGGAAAAACCAAGAGAGAGAGAAAAA 3699
750 euGlyArgSerIleMetArgIleThrGluIleProGluThrValSer 766
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
3700 AAAAACGGAGCAAAACGCGAGAAAAACGAG.....GAACAAAGAAAA 3743
767 Arg.GlnCysSerLysGluAspLysGluTyAlaAspHisGlyThrAlaL 783
    : : : : : : : : : : : : : : : : : : : : :
3744 AACCATAAAAGGAAAAAGCAAAACCGGACAGACCAACAAAAA 3793
783 ysglyThrAlaLeuIleArgLysAsnProGluSerSerGlyAsnThr 799
    ||||| ||||| : : : : : : : : : : : : : : :
3794 AAGGCAGAGAGAAAGCAAAAGAAAAACCAACAAAGAGAGACGAA 3843
800 GlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuLysLy 816
    ||| : : : : : ||||| ||||| ||||| ||||| |||||
3844 AAGAAAAACAGGAAGAGCGCAACAAAAAACCGCACAAAGAGAGAGAG 3893
816 sSerHisSerThrTyAspHisValArgAsp...GlnThrGluGluSerS 832
    : : : : : : : : : : : : : : : : : : : : :
3894 AAGAGGACAGCAACGAAACCAACCAAGAGACCAACCAAGCAACGCA 3943
832 erSerLeuProThrGluSerGlnGluGluThrGluAsnSerThr 848
    ||||| ||| : : : : : ||||| ||||| ||||| |||||
3944 GCGCAACGCCAGGCACAGAGAGAGAAAAACAAACCAACCAACGCAACG 3993
849 LeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLysGluAspSe 865
    : : : : : : : : : : : : : : : : : : : : :
3994 AGGAAACAGCGCGCGAAAAAATAAAGAGAGAGAGAGAGAGAGAGAG 4043
865 rGluAlaGluSer..ThrGluSerValProLeuValCysLysSerAlaSer 881
    ||||| : : : : : ||||| : : : : :
4044 CCAGAGCAGACAAACCAACCAACCAACCAACCAACCAACCAACCAAC 4066
882 AlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMe 898
    ||| : : : : : ||||| ||||| ||||| ||||| |||||
4067 ...AAGAACAGCAGAGAGAGGCAAAACCAACCAACCAACCAACCAAC 4113
898 tLeuGln.LysSerSerValIleAlaSerAlaLysGluLysThrLeu 914
    ||| ||||| : : : : : ||||| ||||| ||||| |||||

```

```

4114 GCGCAAAAAGCCAGCAGC.....AGGAAGCCAAAGAAAGACC... 4154
915 GlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSe 931
   ||| :|:|:| ||| :|:|:|
4155 GGAACCAAAACCAAAACCAACAGAGGCA.....AG 4186
931 rGlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnS 948
   ||| :|:|:| ||| :|:|:|
4187 CAAGCAACCCAGCCACACAGACAGAGGCAAGGAAAGGAAACGCGCGGAAA 4236
948 eAspAsnThrGluThrLys.....AspProAlaProGlnAsn 960
   |||:|:|:| :|:|:|:| ||| :|:|:|
4237 AGCACCACCAAAACCAAAACCAACAGAGAGGAAAGGCGCCAGCGACA 4286
961 SerAsnProAlaGluProArgLysProGlnLysSerGlyLeuMetLys 977
   |||:|:|:| |||:|:|:| |||:|:|:|
4287 AAACAGCCGAGCAAAACCCAGAGAAACAAAGGAAAGGAAAGGCGGAAA 4336
977 sGlnGlnArgValAsnProThrThrAlaAsnSerAspLeuAsnProGlyT 994
   |||:|:|:|:|:|:|
4337 AAGACACAAAGGCCAACAAAGA..... 4358
994 hrThrGlnMetLysAspAsnPheAspIleGlyGluValCysProTrpGlu 1010
4358 ..... 4358
1011 ValTyrAspLeuThrProGlyProValProSerGluSerLysValGlnLys 1027
   |||:|:|:|:|:|:| |||:|:|:|
4359 .....AAGAGAAACCCGCGGAAA 4378
1027 sHisValSerIleValAlaSerGluMetGluLysAsnProThrPheSerL 1044
   |||:|:|:|:|:|:|
4379 AAGA.....AGAAAAAAGCAAAAC..... 4400
1044 euLysGluLysSerHisLysLysProLysAlaAlaGluValCysGlnGln 1060
   |||:|:|:|:|:|:| |||:|:|:|
4401 ..AAGGAGAACCCCAACCAAGCAAAACCAAGAGGAGGAGGAGGAA 4448
1061 SerAsnGlnLysArgIleAspLysAlaGluValCysLeuTrpGluSerG 1077
   |||:|:|:|:|:|:| |||:|:|:|
4449 AAGCGAGAACCCAGA...GAAAAAGGGAA..... 4475
1077 nGlyGlnSerIleLeuGluAspGluLysLeuLeuLeuSerLysThrPro 1094
4475 ..... 4475
1094 aLeuProGluArgAlaLysGluGluAsnGlyGlyGlnProArgAlaAla 1110
   |||:|:|:|:|:|:| |||:|:|:|
4476 .....AAAAACAAACAAACAAAGGAGGAGGAAACCAAGCAGGAGGCA 4520
1111 ...AsnValCysAlaGlyGlnSerGluGluLeuProProLysAlaValAl 1126
   |||:|:|:|:|:|:| |||:|:|:|
4521 AACACACAAAAGAGCGGAGAGCGAGCAAGAAACACCCCGA..... 4562
1126 aSerLysThrGluAsnGluAsnLeuAsnGlnIleGlyHisGlnGluLysL 1143
   |||:|:|:|:|:|:| |||:|:|:|
4563 .AAAAACAAACAAACAAACAAACCAACCAACCAACCAACCAACCAACCA 4611
1143 ysThrSerSerGluGluAsnValArgGlySerTyrAsnSerSerAsn 1159
   |||:|:|:|:|:|:| |||:|:|:|
4612 ACACA...ACGAAAACAAACAAACGAGGAGGAGGAGGAGGAGGAGGAG 4658
1160 AsnPheGlnGlnProLeuThrSerArg 1168
   ||| :|:|:|:| ||| :|:|:|
4659 AAC...AGACAAACCAACAGACGAGGAGG 4682

```

```

seq_name: /SIDS1/gcdata/geneseq/geneseq/NA2000.DAT:AAZ29423

```

```

seq_documentation_block:
ID_AAZ29423 standard; cDNA; 5459 BP.

```

```

XX
AC
AC
XX

```

```

DT 14-MAR-2000 (first entry)
XX
DE Rat gb2 GABA B receptor encoding cDNA.
XX
KW gb2 GABA B receptor; Rat gb2; Rgb2; calcium-sensing receptor;
KW gamma-amino butyric acid; 4-amino butanoic acid; GABA;
KW metabotropic glutamate receptor; neurological disorder;
KW psychiatric disorder; agonist; antagonist; rat; ds.
XX
OS Rattus sp.
XX
Key Location/Qualifiers
FT CDS 151..2973
FT /tag= a
FT /product= "Rat gb2 GABA B receptor"
XX
PN WO9961606-A1.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11869.
XX
PR 29-MAY-1998; 98US-0087274.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Clark J, Bonner TI;
XX
DR WPI; 2000-105616/09.
DR P-PSDB; AAY44343.
XX
PT New GABAB (c-aminobutyric acid or 4-aminobutanoic acid) receptor,
PT useful for identification of (ant)agonists and for treatment of
PT neurological disorders
XX
PS Claim 6; Page 24-25; 67pp; English.
XX
CC The present sequence is a cDNA obtained from rat cortex cDNA library.
CC It encodes gb2 subunit of GABA B receptor. Rat gb2 (Rgp2) shares
CC sequence homology with rat GABA B receptor rglb1, parathyroid cell
CC calcium-sensing receptor and metabotropic glutamate receptors. It can be
CC produced in host cells by transforming them with recombinant expression
CC vector comprising rgb2 encoding cDNA. The cells that express the receptor
CC are used in the development of drugs for treatment of neurological and
CC psychiatric disorders, for pharmacological, physiological, functional, or
CC other investigational analysis of gb2 GABA B receptor, its agonists or
CC antagonists and for determining the ability of a chemical to bind to a
CC mammalian gb2 GABA B receptor in vitro. They may also be used for the
CC preparation of antibodies to rgb2 which can be used in diagnostic assays.
XX
SQ Sequence 5459 BP; 1334 A; 1603 C; 1395 G; 1120 T; 7 other;

alignment_scores:
Quality: 204.00 Length: 1236
Ratio: 0.383 Gaps: 58
Percent Similarity: 43.123 Percent Identity: 19.741

alignment_block:
US-09-775-181-2 x AAZ29423 ..
Align seg 1/1 to: AAZ29423 from: 1 to: 5459
7 proLeuLeuLeuCysLeuLeuAlaGlnLeuGlyLeuGlyAlaValG 23
| | | | | | | | | | | | | | | | | | | | | | | | | |
223 CCCCTGCTGCTGCTGCTGCTG...TGGTTGGCCGCCCGGGCTGGGG 269
23 yAlaSerArgAsp.ProGlnGlyArgProAspSerProArgGlu..... 37
| :|:|:| ||| :|:|:|
270 CTGGAGCGGGGGCGCCCGCGCGCCCGCCAGCAGCGCGCGCTGCCA 319
38 .....ArgThrProLysGlyLysProHisAlaGlnGln... 48
| | | | | | | | | | | | | | | | | | | | | | | | | |

```

320 TCATGGCGCTCATGCCGCTCACCAGGAGGTGGCCAAAGGGCAGCATCGGG 369

49ProGlyArgAlaSerAlaSerA 56
::: |||||:::

370 CGCGCGGTCTCCCGCGCGTGGAGCTACCCATCAGCAGATCCGCAACGA 419
::: |||||:::

56 sPserSer..AlaPro..... 60
:: ||| |||||

420 GTCACCTCGTGGCGCCCTACTTCTTGGACCTCGCACTCTATGACACCGAGT 469
::: |||||:::

61TrpS 62
::: |||||:::

470 GTGACATGCAAAGGAGCTGAAGCCCTCTATGACGCAATAAGATATGGG 519
::: |||||:::

62 erArgSerThrAspGlyThrIleLeuAlaGlnLysLeuAlaGluGluVal 78
::: |||||:::

520 CGCAACCATTTGATGGTCTTTGGAGGCGTCGTCCGCTCTGTCACTAT 569
::: |||||:::

79 ProMetAspValAlaSerTyLeuTyThrGlyAspSerHisGlnLeuLy 95
:: ||| ::||| :::

570 TATCGCGGAGTCCCTCCAAGGCTGGAATCTGTCAGCTTCTCTCCGCG 619
::: |||||:::

95 sArgAlaAsnCysSer..... 100
::: |||||:::

620 CCACACAGCGCTCTCTCGGATAGAGAAGTACCCGATATTTCTCCGG 669
::: |||||:::

101 ..GlyArgTyGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAla 116
::: ||| ::||| ::|||

670 ACGGTCCGTCAGACAACCGGTGAACCC..... 698
::: |||||:::

117 SerAlaHisProSerLeuHisArgAlaLeuAspThrLeu..... 129
::: |||||:::

699 ...CGCCATCTCTGAAGCTCCTGAAGCACTTCCGCTGGCGGTCTGGCCA 745
::: |||||:::

130 ThrHisAla..... 132
::: |||||:::

746 CACTCACGACGAGCYGACGCCCTCTCCGAGGTGAGGAATGACCTGACT 795
::: |||||:::

133ThrAspPheLeu 136
::: |||||:::

796 GGGTTCCTATGGGAAGATATTGAGATCTCACACACAGAGTTCTC 845
::: |||||:::

137 AsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeuAsnAsp 153
::: |||||:::

846 CAATGATCCCTGCACAGCGTCAAAAAGCTCAAGGGGAATGACGTGGCGA 895
::: |||||:::

153 pIeuAspTrp..TyrClnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
:: ||| ::||| ::|||

896 TCATCTTGGCCGAGTTGACCAAGATATGGCAAAAGTCTTCTGTGT 945
::: |||||:::

169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuSerAlaP 186
::: |||||:::

946 GCCTTCGAGGAGAGCATGTTTGGCAGCAAGTACCAGTGGATCATCCCGG 995
::: |||||:::

186 roAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArgIleLeu 202
::: |||||:::

996 ATGTTACAGCGCTG.....CGTGGTGGGACGAGGTGCATG 1030
::: |||||:::

203 LeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrLeuGl 219
::: |||||:::

1031 TGGAGGCCAATT.....CCTCACGCTGCCGCA.....GA 1062
::: |||||:::

219 uThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHis 235
::: |||||:::

1063 ACCCTCTTGGCTGCCATGGA.....AGGTTACAT 1091
::: |||||:::

236 ArgArgGly.....ProAsnGlnGlyProArgGlyLeuGlyHis 249
::: |||||:::

1092 CGGAGTGGACCTTGAGCCCTCAG..... 1115
::: |||||:::

249 rTrpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLys..Trp 265
::: |||||:::

1116CTCCAACAACAAATCAAGACACTCTCAGGGAAG 1146
::: |||||:::

[illegible]

```
525 ArgValMetArgMetLeuAlaValIleLeuLeuValValPheTrpPheLeu 541
1957 .....ATGCTGCTCATGACACCTGTGCTATCT 1982
541 uileGlyTrpThrSer..... 546
1983 GATCTGTGGCAGGCTGTGGACCCCTGCGGAGGACAGTGGAGAGGTACA 2032
546 ..... 546
2033 GCATGGAGCGGACCCAGCAGCGCGGACATCTCCATCCGCCCATTTGCTG 2082
547 SerValCysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLy 563
2083 GAACACTGGGAAACACCCACATGACATC..... 2112
563 sThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspT 580
2113 .....TGGCTTG 2119
580 yrMetThrAlaValAlaGluPheLeuPheLeuLeuTrpGlyValTyrLeu 596
2120 GCATTGCTACGCTACAGGGCTCCTCATGCTATTGCTGTTCTTG 2169
597 CysTyrAlaValArgThrValPro....SerAlaPheHisGluProArgTyr 612
2170 GCATGGAAACCCGCAATGTGACATCCCTGCGCTCAACGACAGCAAGTA 2219
612 rMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePheHisT 629
2220 CATCGGCATGAGTGTACAATGTGGGGATCATGTCCATCATCGGGGCTG 2269
629 hrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrp..... 642
2270 CTGTCCTCTCTGACGCGTGACACGCCCAACGTCGATCTGTCATCGTG 2319
643 MetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrI 659
2320 GCCTGTGTCATCTCTTGCAGC.....ACCATCACTCT 2354
659 eGlyLeuLeuLeuIleProLysPheSerHisSerSerAsnAsnProArgA 676
2355 CTGCTGCTGTTGTGCCAAGCTCATCACTCTGAGGACAAACCCCT.... 2400
676 sPAspIleAlaThrGlu.....AlaTyrGluAspGluLeuAspMetGly 690
2401 ..GAGCAGCCACTCAGACAGCGGTTCCAGTTCACACAGAACACAGAG 2448
691 ArgSerGlySerTyrLeuAsnSerSerIleAsnSerAlaTrpSerGluHi 707
2449 AAGAAGATTCGAAGACCTCCACTTCAGTCACACGCGTGAACGCGGAG 2498
707 sSerLeuAspProGluAspIleArgAspGluLeuLysLeuTyrAlaG 724
2499 CAGCTCACCCCTGGAGGACTCGAGTCAGAACCAACCCGCTT..... 2541
724 InLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnProHisLeu 740
2542 .....CGAATGAAGATCACAGAGCTGGAGAACAGACTTG 2574
741 GlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgA 757
2575 GAAGAAGTCACCATCAGTACAG.....ACACACACAGAGAA 2612
757 rgIleThrGluIleProGluThrValSerArgGlnCysSerLysGluAsp 773
2613 GA.....CCACATACATCAACAGAAATCACTACCAAGAGCTCA 2650
774 LysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuIleArgLy 790
2651 ACGACATCCTCAGCTGGCACTTCACAG...AGAGCACAGATGGAGGA 2697
790 sAsnPro.....ProGluSerSerGlyA 798
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ50616
seq_documentation_block:
ID AAZ50616 standard; cDNA; 2922 BP.
XX
AC AAZ50616;
```



```

XX 20-JUN-2000 (first entry)
XX
XX Human calcium sensing receptor isoform, CaSRd encoding cDNA.
XX
XX Calcium Sensing Receptor; CaSRd; isoform; human; splice variant;
XX hypotensive; osteopathic; receptor activity; calcium level; modulator;
XX treatment; hyperparathyroidism; osteoporosis; Paget's disease;
XX hypercalcaemia malignancy; hypertension; gene therapy; ss.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..2922
XX /tag= a
XX /product= "Human calcium sensing receptor isoform, CaSRd"
XX /note= "Does not include stop codon"
XX /partial
XX
XX WO200006601-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17116.
XX
XX 30-JUL-1998; 98US-0094702.
XX
XX (AVET ) AVENTIS PHARM PROD INC.
XX
XX Yu KT, Labaudiniere RF, Thrower LW;
XX
XX WPT; 2000-195263/17.
XX
XX P-PSDB; AAY45000.
XX
XX Nucleic acids encoding isoforms of human calcium sensing receptor for
XX treating, e.g. hyperparathyroidism or osteoporosis
XX
XX Claim 8; Page 60-66; 81pp; English.
XX
XX The present sequence is the cDNA encoding an isoform of the human
XX calcium sensing receptor CaSRd, expressed in the kidney. It is a splice
XX variant of the wild type CaSRd, that arise from deletion of nucleotides
XX 1080-1391, from the extracellular domain, that comprise acidic residues.
XX It has hypotensive and osteopathic activity. The CaSR isoforms can be
XX used to identify agonists and antagonists that modulate the receptor
XX activity and calcium levels. These modulators are useful for treating
XX hyperparathyroidism or osteoporosis. Paget's disease, hypercalcaemia
XX malignancy or hypertension. The DNA sequence is also useful for altering
XX the CaSR activity and in gene therapy.
XX Note: This sequence has deletion of nucleotides 1080-1391, according to
XX the sequence shown in the specification, but has been stated as
XX 1075-1386 in the claims.
XX
XX Sequence 2922 BP; 555 A; 848 C; 752 G; 667 T; 0 other;

```

```

alignment_scores:
  Quality: 200.50      Length: 941
  Ratio: 0.466        Gaps: 40
  Percent Similarity: 45.696  Percent Identity: 19.766

```

```
alignment_block:
```

```
US-09-775-181-2 x AAZ50616
```

```
Align seg 1/1 to: AAZ50616 from: 1 to: 2922
```

```

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   ::::::::::::::::::::
394 TCAGAGCACATTCCTCTACGATTGCTGGTGGAGCAACTGGCTCAGG 443
   ::::::::::::::::::::
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :   :   :   :   :   :   :   :   :   :   :   :   :

```

```

444 CGTCCTCCAGCGCAGTGGCAAAATCTGCTGGGGCTCTTCTACATT..... 486
108 euProGlyLysTyrProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   ||| ::::: |||
487 .....CCCCAGGTCAGTTATGCTCTCCAGC.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   ||| ::::: ::::: ::::: ::::: :::::
517 CTCCTCAGCAACAAGATCAATTCAGTCTTCTCTCCGAACC...ATCCC 563
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTyrG 158
   ::::: ::::: ::::: ||| ::::: :::::
564 CAATGATGAGCAGCAGCCACTGCCATGCCAGACATCATCGAGTATTTCC 613
158 InAlaLeuValTyrSerLeuGluGluProSerIleSer..... 172
   : : : : : ||| :::::
514 GCTGGAACCTGGTGGGCACAATTGACAGCTGATGACGACTATGGCGGCGC 663
173 .....ArgAlaAlaIleThrPheSerT 180
   ||| ::::: |||
564 GGGATTGAGAAATCCGAGAGGAAGCTGAGGAAGGGATATCTGCATCGA 713
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   ||| ::::: |||
714 CTTCAGTGAATCATCTCCAGTACTCTGATGAGGAGAGATCCAGCATG 763
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   ||| ::::: |||
764 TGGTAGAGGTGATTCAAAATTCACGGCCAAAGTCATCGTGTTCCTCC 813
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   ||| ||| ||| ||| ||| :::::
814 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATTGTCGGCGCA... 860
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   ::::: ||| ::::: |||
861 .....TATCACGGCAAGATCTGGCTGGCAGCGAGCCCTGGGCCAGCT 904
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   ||| ::::: |||
905 CCTCCTGATGCCATGCTCAGTACTTCCAGTGGTGGCGGCACCAT 954
250 rpArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
   ||| ::::: |||
955 GGATTGCTCTGAAGC.....TGGGCA 977
267 ProProTyrLeuGlu.....Cy 272
   ||| |||
978 GATCCAGCGCTTCGGGGAATTCCTGAGAGAGTCCATCCAGGAGTCTG 1027
272 s.GluAsnGlySerTyrLysProGlyTyrLeuValThrLeuSerSerAla 288
   : : : : : |||
1028 TCCACAATGGTTTCCCAAGGAGTTTGGGAAGAAACATTTAACTGCCAC 1077
289 IleTyrGlyLeuGlnProAsnLeuValProGluPheArgGlyValMetLy 305
   : : : : : |||
1078 CTACGGCATCTA.....AACTTTACAAACAATATGGG 1109
305 sValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSerAspG 322
   ::::: ||| ::::: |||
1110 G.....GAGCAGGTGACCTTTGATGAGTGT.....G 1135
322 LyTrpPheSerGlyThrHisLysCys.....HisLeuAsnAsnSer 335
   ||| ::::: |||
1136 GTGACCTGGTGGGNACTATTCCATCATCACTGGCACCTCTCCACAG 1185
336 Glu...CysMetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrG1 351
   : : : : : |||
1186 GATGGCTCCATCGTGTTTAAGGAAGTCGGGTAT.....TACAA 1223
351 uCysIleCysLysAlaGly.....PheTyrHisProGlyValLeuP 365
   : : : : : |||
1224 CGTCTATCCCAAGAGGGAAGAAAGACTCTTTCATCAACGAGGAGAAATCC 1273

```


||||:| :|||
2882 GTGGTGGAGGAGCACT 2898

seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AA86167

seq_documentation_block:

ID AA86167 standard; DNA; 3219 BP.

XX AC AA86167;

XX DT 03-DEC-1997 (first entry)

XX DE Nucleotide sequence of pratch3.

XX KW Chimeric receptor; extracellular domain; seven transmembrane domain;
intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;
neurological disease; ss.

XX OS Chimeric - Rattus rattus.

XX FH Location/Qualifiers
FT 25..3195
CDS /*tag= a
FT /product= pratch3
FT

XX MN W09705252-A2.

XX PD 13-FEB-1997.

XX PF 25-JUL-1996; 96WO-US12336.

XX PR 26-JUL-1995; 95US-0001526.

XX PA (NPSP-) NPS PHARM INC.

XX PI Fuller FH, Hammerland LG, Krapcho KJ;

XX DR WPI; 1997-145690/13.

XX DR P-PSDB; AAW25764.

XX PT Chimeric receptors comprising metabotropic glutamate receptor and
calcium receptor - used for screening for neurologically active
compounds

XX PS Example 4; Fig 4; 177pp; English.

XX CC This sequence encodes a chimeric receptor. The chimeric receptor
comprises an extracellular domain, a seven transmembrane domain, and
an intracellular cytoplasmic tail domain, and a sequence of at least
6 contiguous amino acids is homologous to a sequence of a metabotropic
glutamate receptor (MGR), and a sequence of at least 6 contiguous amino
acids is homologous to a sequence of a calcium receptor (CR). The
chimeric receptor may be used for screening for compounds that bind to
or modulate the activity of MGR or CR and for determining the site-of-
action of a CR active compound. The compounds can be used in the
treatment of neurological diseases and disorders. They can also be
used as diagnostic agents. Chimeric receptors such as this, allow the
coupling of certain functional aspects of an MGR with certain functional
aspects of a CR. They allow for more efficient high-throughput screening
of compounds.

XX SQ Sequence 3219 BP; 775 A; 880 C; 851 G; 713 T; 0 other;

alignment_scores:

Quality: 199.50 Length: 964
Ratio: 0.503 Gaps: 43
Percent Similarity: 41.183 Percent Identity: 20.228

alignment_block:

US-09-775-181-2 x AA86167 ..

Align seg 1/1 to: AA86167 from: 1 to: 3219

```

153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
501 GATGGCCCTGGCTCCAGCTCTGTGGCCATTCAGTCCAGATCTTCTCC 550
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
551 A.....GCTGTTCGACATCCACAGATCGCTATTCCTGCC 585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 ACAAGCATAGACCTGAGTGACAAACTTTGACAAATACCTTCCTGAGGT 635
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 .....GTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 GGTCCCTTCTGACACTTGTGAGGCAAGGCGATGCTGCACATACCTCAAGC 685
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 GTTACAACTGGACCTATGCTCTCAGCAGTCCACACAGAGGAA..... 728
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
729 .....TTACGGGAGAGTGAATGGATGGATGCTTT 755
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
756 CAAAGAACTGGCTGCCCGAGGAGGCT...CTGCATCGCACACTCGGACA 802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 OAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
803 AAATCTACAGCAATGCTGGGAGAGAGCTTTGACGGGCTCCTCGCTAAA 852
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 .....GloglyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 CTCGGGAGGGCTTCCCAAGGC...CAGGGTTGTGTCTCTCTCTCGCA 899
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ArgLysAspGlyLeuGlyAspLysSerHis..PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
900 GGGCATCACAGTGGGGCTTACTGAGTGCATCGCGCGCTGGCGCTCG 949
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....SerProTyrLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
950 TGGCGAGTTCCTCATTGGAAGTGTGATGGATGGCGAGACAGATGAA 999
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1000 GTCATCGAAGGCTATGAGTGGAAAGCCACGGAGGATCAATAAGCT 1049
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1050 TCAGTCTCCAGAGTCAAGTCAATTTGATGACTACTTCTTGAAGCTGAGC 1099
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1100 TGGACACCAACACAAGGAATCCTTGGTTCCCTGAGTTCTGGCAACATCGC 1149
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu..... 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1150 TTCAGGTGTCGCTTACCTGACCTCTTGGAAACCCCACTTTAAGAA 1199
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1200 AGTGTGCACAGGAATGAAGCTTGGAAAGAAACTATCTCCAGGACAGCA 1249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrH1 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1250 AAATGGGATTTGTCAATCAATGCCATCTATGCCATGGCACATGGCTCGAG 1299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 sLysCys..His.....LeuAsnAsnSerGluCysMe 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1300 AACATGCACCATGCTCTGTGTCCCGGCAATGTGGGCTGTGTGTATGCTAT 1349
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

3033 GACCGTCCAGGAACAGTCTGCAAGGACCTGTGGTGGAGACCAGCGG. 3081
      ::||| |||::: ||| ::: ::::
817  erHisSerThrTyrAspHisValArgAspGlnThrGluGluSerSerSer 833
      ::|||::: ||| ||| :::
3082 .....CCAGAGGTGGAGGACCCCTGAAGAGTTGTCCCCACCA 3117
      ::|||::: ||| ||| :::
834  LeuProThrGluSerGlnGluGlu.....GluTh 843
      ||| ::::|||::: ||| :::
3118 CTTGTAGTGTCCAGTTCACAGAGCTTTGTCATCAGTGGTGGAGGAGCAC 3167
      ||::: ::::|||::: ||| :::
843  rThrGluAsnSerThrLeuGluSerLeuSerGlyLysLys 856
      ||::: ::::|||::: ||| :::
3168 TGTTACAGAAAACGTAGTGAATTCATAAATGAAGGAGA 3207

```


91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
: : : : :
484 CGTCTCCAGCGGAGTGGCAAACTGCTGGGGCTCTTCTACATT 526
108 euProGlyLysrppProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
: : : : :
527CCCGAGGTCAGTATGCTCCCTCCAGCAGA 556
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
: : : : :
557 CTCTCAGCAACAGNATCAATCAAGTCTTCTCCCGAACCCATCCC 503
141 nSerAsnLysSerArgGluGlnAnLysLeuGlnAspLeuAspTrpTyrG 158
: : : : :
604 CAATGATGAGCACCGGCGCACTGCCATGGCAGACATCATCGAGTATTTC 653
158 lnAlaLeuValTrpSerLeuGluGlyGluProSerIleSer 172
654 GCTGGAACCTGGGTGGGCAATATGACGCTGTAGCAGCTATGGGCGCGG 703
173ArgAlaAlaIleThrPheSerT 180
704 GGGATTGAGAAATTCGAGAGAACTGAGGAAGGGATATCTGCATCGA 753
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
: : : : :
754 CTTCACTGAACCTCATCTCCAGTACTCTGATCAGGAAGAGATCCAGCATG 803
196ArgGluGluSerArgIleLeuLeuG 204
804 TGGTAGAGGTGATCAAAATTCACGGCCCAAGTATGCTGTTTCTCC 853
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
: : : : :
854 ACTGCCAGATCTTGAGCCCTCATCAGGAGATGTCGCGCGCAA... 900
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237
: : : : :
901TATCAGGGCAAGATCTGCTGGCCAGCGAGGCTGGGCCAGCT 944
237 gGlyProAsnGlnGlyProArgGlyLeuGlyHisSerT 250
: : : : :
945 CTTCCCTGATGCCAGTCTAGTACTTCCAGTGGTGGCGGCACTT 994
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
: : : : :
995 GGATTGCTCTGAAGGCTGGGCA 1017
267 ProProTyrLeuGluCy 272
: : : : :
1018 GATCCAGGCTTCCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1067
272 sGluAsnGlySerTyrLysProGlyTrpLeuValThr 284
: : : : :
1068 TCACAAATGGTTTCCCAAGGAGTTTGGGAAGAACAATTTAACTGCCAC 1117
285 LeuSerSerAlaIleTyrGlyLeuGlnPro 294
: : : : :
1118 CTCCAAGAGGTGCAAAAGGACCTTACTCTGGACACCTTCTGAGAGG 1167
295AsnLeuValProGluPheArgg 302
: : : : :
1168 TCACGAAGAAGTGGCGACAGGTTTAGCAACAGCTCCGACAGCTTCCGAC 1217
302 lyValMetLysValAspIleAsnLeuGlnLysVal 313
: : : : :
1218 CCCTCTGACAGGGGATGAGAACAATCAGCAGTGTGCGAGACCCCTTACATA 1267
313 313
1268 GATTACAGCATTTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1317
314AspIleAspGlnCys 318
: : : : :
1318 CATTCGCCACGCTTTCGAAGATATATACCTTACCTGGGAGAGGCG 1367
319SerSerAspGlyTrpPhe 324
: : : : :
1368 TCTTCACCAATGGCTCCTGTGCAGACATCAAGAAAAGTTGAGCGGTGGCAG 1417
325 SerGlyThrHisLysCysHisLeuAsnAsnSer 335
: : : : :
1418 GTCTTGAAGCACCTACCGCACTAACTTACAAACAATATATGGGGAGCA 1467
336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
: : : : :
1468 GGTGACCTTTGATGAGTGTGGTGACCTGTGGGGAACATTATCCATCATCA 1517
348Gly 348
: : : : :
1518 ACTGCGACCTCTCCAGAGGATGGCTCCATCGTGTAAAGGAAGTCGGG 1567
349 AlaTyrGluCysIleCysLysAlaGlyPheTyrHisProG 362
: : : : :
1568 TATTACAAGCTATGCCAAGAGGAGAAAGACTCTTCATCATCAACGAGGA 1617
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
: : : : :
1618 GAAATCTCTGGAGTGTCTCCAGGAGGTGCCCTTCTCCAACATGCA 1667
378IleSerGlySerThrLysAspValSerGluGlu..... 388
: : : : :
1668 GCCGAGACTGCTGGCAGGAGGACCAAGAAAGGATCATTTGAGGGGAGGCC 1717
389AlaTyrValCysLeuProCysArgGluGly 398
: : : : :
1718 ACCTGCTGTTGAGTGTGGAGTGTCTGATGGGAGATAGTAGTCA 1767
399CysProPheCysAlaAspAsp 405
: : : : :
1768 GACAGATGCCAGTGCCTGTAACAAGTCCAGCTGCTGTTGTTGTTGAGTGGAGT 1817
406SerProCysPheValGlnGluAspLysTyrLeuArg 417
: : : : :
1818 AGAACACACCTCTGCGAGGCCAATCTGCTGCTGTTATCTTGAAGTGGAGT 1867
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
: : : : :
1868 GACATAGATCTATCATAGCCATCGCTTTCTTCCCTGGGCATCCTCGT 1917
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
: : : : :
1918 GACGCTGTTGTACCCCTCATCTTGGTTCTGTACCGGGACACACCCGTGG 1967
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
: : : : :
1968 TCAATCTCCAGTAGGAGCTGCTATATCATCTCTGGCTGTTGTTTTC 2017
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
: : : : :
2018 CTGCGCTATGTGCCCTTTCACCCCTCATCCCAACACCTACTACACATC 2067
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
: : : : :
2068 CTGCTACCTCCAGGCGCTTCTAGTTGGCTCTCTTCTGCCATGTGCTACT 2117
497 lyThrValThrLeuLysLeuHisArgValLeuLysVal 509
: : : : :
2118 CTGCTTTAGTACCAAAACCAATCGTATTGACGCACTCTGCTGGCTGGCAGC 2167
510PheLeuSerArgThrAlaG 516
: : : : :
2168 AAGAAGAAGATCTGCACCCGGAAGCCAGATTCATGACGCTGGGCCCA 2217
516 nArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaValI 533


```

1100 ysGluGluAsnGlyGlyGlnProArgAlaAalaAsnValCysAlaGlyGln 1116
      ::      ::|||      ::|||      ::|||      ::|||      ::|||
3543 CTGCCCTGACGCTCCTTCTCCTTTCCGAGATTCCGTGGCTCTGGCAGC 3592

      ::|||      ::|||      ::|||      ::|||      ::|||
1117 SerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluAs 1131
      |||      |||      ::|||      ::|||      |||
3593 TCAGTGGCCAGTTCCGCCCTATCTAGTCGGTCTCTGCACCCCTCCAAA 3642

      ::|||      ::|||      ::|||      ::|||      ::|||
1131 nGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSer 1147
      |||      ::|||      ::|||      ::|||      ::|||
3643 TGTAACTTACCGCTCTGTCAATTCTGAGGAGCATACAAAGACTCTTCC 3691

seq_name: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq: PCT-US91-09422-16

```

```

seq_documentation_block:
: Sequence 16, Application PC/TUS9109422
: GENERAL INFORMATION:
: APPLICANT: Mulvihill, Eileen R.
: APPLICANT: Hagen, Frederick S.
: APPLICANT: Houamed, Khaled M.
: APPLICANT: Almers, Wolfhard
: TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend
: STREET: One Market Plaza, Steuart Street Tower
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94105-1492
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US91/09422
: FILING DATE: 19911212
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,007
: FILING DATE: 18-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/648,481
: FILING DATE: 30-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/626,806
: FILING DATE: 12-DEC-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Parmelee, Steven W.
: REGISTRATION NUMBER: 31,990
: REFERENCE/DOCKET NUMBER: 13952-6PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 467-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5236 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: CLONE: SN23
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 627..3344
: PCT-US91-09422-16

```

```
alignment_scores:
  Quality: 237.00
  Length: 1248
```

```

Ratio: 0.479      Gaps: 54
Percent Similarity: 39.663      Percent Identity: 19.872

alignment_block:
US-09-775-181-2 x PCT-US91-09422-16  ..

Align seg 1/1 to: PCT-US91-09422-16 from: 1 to: 5236

153 AspLeuAspTtPtyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1103 GATCGGCCCTGGCTCCAGCTCTGTGGCCATTCAAGTCAGCAATCTTCTCC 1152
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1153 A.....GCTGTTGCGACATCCACAGATCGCCTATTCTGCC 1187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1188 ACAAGCATAGACCTCAGTGCACAAACTTTGTACAAATACTTCTCTGAGGT 1237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 .....GCTGTTGCGACATCCACAGATCGCCTATTCTGCC 1201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1238 GGTCCCTTCTGCACATTTCAGGCAAGGCGATGCTGCACATAGTCAAGC 1287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1288 GTTACAACCTGGACCTATGCTCAGCAGTCCACACAGAAGGAA..... 1330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1331 .....TTACGGCGAGAGTGAATGGATGCTTT 1357
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1358 CAAGAAGTGGCTGCCAGGAAAGGCT...CTGCATCGCACACTCGGACA 1404
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405 AAATCTACAGCAATGCTGGCGAGAAGAGCTTTGACCGGCTCTCGCGTAA 1454
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1455 CTCGGGAGCGGCTTCCCAAGGC...CAGGGTGTGGTCTGCTTCTGGGA 1501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ArgLysAspGlyLeuGlyGlyAspLysSerHis...PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1502 GGGCATGACAGTGGCGGCTTACTGATGGCCATGGCGCGCTGGCGGCTG 1551
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....SerProTrpTrpLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 TGGCGAGTTCTCACTACTTGAAGTGTGATGGTGGCAGACAGAGATGAA 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 GTCATCGAAGGCTATGAGGTGGAAGCAACGAGGAGGATCACAATAAGCT 1651
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 TCAGTCTCAGAGTCAGGTCATTGTGATGACTACTTCTGAAGCTGAGCG 1701
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1702 TGGACCAACACAGGAATCCTTGGTTCCTGAGTTCCTGCGCAACATCGC 1751
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu.... 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1752 TTCAGTGTGGCTTACCTGGACACCTCTTGGAAACCCCACTTTAAGAA 1801
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1802 AGGTGTCACAGGAATGAAGAGCTGGGAAGAACTATGTCAGGACAGCA 1851
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
    |||   |||       |||      |||     |||:::|||
853 GATCGGCCCTGGCTCCAGCTGTGTGGCATTCAAGTCACAGAATCTCTCC 902
    |                                     |||||
169 oSertleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
    -               ||||||           |||         |||
903 A.....GCTGTTCGCATCCCACACAGATCGCATTTACTTGCC 937
    |||          |||        |||       |||:::|||
184 .....SerAlaProAlaProGlnValPheLeuAlaThrArgGlu... 197
    |||          |||        |||       |||:::|||
938 ACAAGCATAGACCCTGAGTGACAANAACCTTTGACAAATACTTCCTGAGGGT 987
    |||          |||        |||       |||:::|||
198 .....GGTCCCTTCTGACACTTTTCACGAAGGGCGATGCTCGACATAGTCAGC 1037
    |||          |||        |||       |||:::|||
201 eLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrL 218
    |||          |||        |||       |||:::|||
1038 GTTAACAACCTGGACCTATGCTCTCACGAGCTCCACACAGAAGGAA.... 1080
    |||          |||        |||       |||:::|||
218 euGluthrCluTrpPheHisGlyLeuArgAsgLysTrp..... 230
    |||          |||        |||       |||:::|||
1081 .....TTACGGCGAGAGTAGGAATGGATGCGTT 1107
    |||          |||        |||       |||:::|||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
    |||          |||        |||       |||:::|||
1108 CAAAGAACCTGGCTGCCACGGAAGSGCT..CTGCATCGCACACTCGGACA 1154
    |||          |||        |||       |||:::|||
239 oAasn..... 240
    |||          |||        |||       |||:::|||
1155 AAATCTACGCAATGCTGGCAGAGAAGAGCTTTGACCGGCTCCTCGGTAAA 1204
    |||          |||        |||       |||:::|||
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
    |||          |||        |||       |||:::|||
1205 CTCCTGGGAGCGGCTTCCCCAAGGC...CAGGGTGTGGTGCTGCTTCGCGA 1251
    |||          |||        |||       |||:::|||
252 ArgLysAspGlyLeuGlyAspLysSerHis..PheLysTrp..... 265
    |||          |||        |||       |||:::|||
1252 GGCGATGACAGTCGCGGCTTACTGAGTCGCATCGCGCGCTGGCGCTCG 1301
    |||          |||        |||       |||:::|||
266 .....SerProProTyrlLeuGlcysGluAsnGlyser..... 276
    |||          |||        |||       |||:::|||
1302 TGGCGGAGTCTCTCACTCATTTGGAAGTGATGGATGGCAGACAGAGATGAA 1351
    |||          |||        |||       |||:::|||
277 .....TyrlLysPro..... 279
    |||          |||        |||       |||:::|||
1352 GTCATCGAAGGCTATGAGGTGGAGCCCAAGGAGGATCACAAATAAGCT 1401
    |||          |||        |||       |||:::|||
280 .....Gly 280
    |||          |||        |||       |||:::|||
1402 TCAGTCTCCAGAGGTCAGGTCAITTGATGACTACTTCTGAGCTGAGGC 1451
    |||          |||        |||       |||:::|||
281 TrpLeuValThr.....LeuSe 286
    |||          |||        |||       |||:::|||
1452 TGGACACCAAACAAGAAGATCTTGTTTCCCTGAGTTCTTGCCAACATCGC 1501
    |||          |||        |||       |||:::|||
286 rSerAlaIleltyrGlyLeuGln.....ProAsnLeu.... 296
    |||          |||        |||       |||:::|||
1502 TTCCAGTCTGCCCTACCTGGACACTCTTTGAAACCCCACTTAAAGAA 1551
```

```

2443 CAAGAAGAAGATCTGCACCCGGAAGCCAGATTTCATGAGCGCTGGGCC 2492
||||:|||||
516 InArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
||| |||:||||:|||||
2493 AAGTGATCATCCCTCCATTCGATTAGTACAGCTAACACTAGTGGTG 2542
||| |||:||||:|||||
533 IleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerValCy 549
||||:|||||
2543 ACCTTGATCATC 2554
549 sGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSera 566
||||:|||||
2555 .....ATGAGACCTCCCATCCCATTTTGTCTACCCGAGTATCAAGG 2597
||| |||:||||:|||||
566 sphHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
||| |||:||||:|||||
2598 AA.....GTCTACCTTATCTGCATACCAAGCACTGGGTAGTGCC 2641
||| |||:||||:|||||
583 AlaValAla...GluPheLeuPheLeuLeuTrpGlyValTyrLeuCysTy 598
||||:||||:||||:|||||
2642 CCTGTGGTTACAATGGAGCTCTCATCATGAGCTGTACCTACTATGCCT 2691
||| |||:||||:|||||
598 rAlaValArgThrValProSerAlaPheHisGluProArgTyrMetAlav 615
||||:||||:||||:|||||
2692 CAAGACCCGCAACGTCGCGGCAACTTCAATGAGGCTAAATACATCGCCT 2741
||| |||:||||:|||||
615 alAlaValHisAsnGluLeuIleIleSerAlaIlePheHisThrIleArg 631
||||:||||:|||||
2742 TCACCATGTACACTACCTGCATCATCTGGCTGGCTTTCGTCCCATTTAC 2791
||| |||:||||:|||||
632 PheValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPh 648
||| |||:||||:|||||
2792 TTT.....GGGAGCAACTACAAGATCATCATACCTG 2823
||| |||:||||:|||||
648 eaIaHisThrHisLeuThrValThrValThrIleGlyLeuLeuIleP 665
||||:||||:||||:|||||
2824 CTTCCGGGTGACCTTCAGTGTGACCGTGGCTGGGTGCATGTTTACTC 2873
||| |||:||||:|||||
665 rOLys.....PheSerHisSerSerAsnAsnProArgAsp 677
||||:||||:|||||
2874 CGAAGATGTACATCATCTGCCAACCTGAGAGGAAGCTCCGCAAGTCC 2923
||| |||:||||:|||||
678 IleAlaThrGluAlaTyrGluAspGluLeuAspMetGly..... 690
||||:||||:||||:|||||
2924 TTCAGGACC.....TCTGATGTTTCCGCATGCACGTCGGTGATGG 2964
||| |||:||||:|||||
691 .....ArgSerGlySerTyrLeuAsnSerSerIleAsnSera 703
||||:||||:||||:|||||
2965 CAAACTGGCGTCCGCTCCACACCTTCTCTCAAC..... 2998
||| |||:||||:|||||
703 laTrpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLys 719
||| |||:||||:|||||
2998 ..... 2998
720 LysLeuTyrAlaGlnLeuGluIleTyrLysArgLysLysMetIleThrAs 736
||||:||||:||||:|||||
2999 .....ATTTTCGGAGAAAGACCCCGGGCAGG 3027
||| |||:||||:|||||
736 naSnProHisLeuGlnLysLysArgCysSer..... 746
||||:||||:||||:|||||
3028 GAATCCCAATTCATCGGCAAGTCTGTCTCATGTGGTCTGAACAGGTGGAA 3077
||| |||:||||:|||||
747 ..LysLysGlyLeuGlyArgSerIleMetArgArgIleThrGluIlePro 762
||| |||:||||:||||:|||||
3078 GACAGCGCCCAAGGACAGCAGCTGTGGCAGCGCTCTCTCTGCACGTG 3127
||| |||:||||:|||||
763 GluThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHi 779
||||:||||:||||:|||||
3128 AAGACCAACGAGACGCGCTGTAAACAA..... 3154
||| |||:||||:|||||
779 sGlyThrAlaLysGlyThrAlaLeuIleArgLysAsnProProGluSers 796
||||:||||:||||:|||||

```

```

3155 .....ACAGCCGTAAATCAAA.....CCCTCACTA 3179
796 erGlyAsnThrGlyLysSerLysGluGluThrLeuLysAsnArgValPhe 812
||||:||||:||||:|||||
3180 AAAGTTACCAAGCTCTGGCAGAGCCCTGACCTTTTCAGAT..... 3220
813 SerLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGl 829
3220 ..... 3220
829 uGluSerSerSerLeuProThrGluSerGlnGluGluThrThrGluA 846
3220 ..... 3220
846 snSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLeuLys 862
3221 .....GCCAGCACCAAGACCTTTTACNATGTGGAA 3250
863 GluAspSerGluAlaGluSerThrGluSerValProLeuValCysLysSe 879
||||:||||:||||:|||||
3251 GAAGAGGACAATACCCCTTCTGCTCCTCAGCCCTCCAGCAGCCCTTC 3300
879 rAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgT 896
||| |||:||||:||||:|||||
3301 TATGTGTGTGCACCGACGCGGGCCACCGTGGCCACACACCTCTGC 3350
896 hrSerMetLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLys 912
||| |||:||||:||||:|||||
3351 CACCCCATCTG.....ACCGAGAAGAGACC 3376
913 ThrLeuGlyLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgTh 929
||| |||:||||:||||:|||||
3377 CCCCTGTCTCTGGCTGATTCCGTATCCCAAGGGCTTG..... 3415
929 rLysSerGlnLysProLeuProLysAspLysGluThrAsnArgAsnHis 946
||||:||||:||||:|||||
3416 .....CCTCTCTCTCCCGCAGCAGCAG..... 3439
946 erAsnSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsn 962
3440 .....CCACAGCAGCGCGCC 3454
963 ProAlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGl 979
||| |||:||||:||||:|||||
3455 CCTCAGCAGCGCGCGCAGCAGCAG.....TCCCTGATGGACCACT 3498
979 nArgValAsnProThrThrAlaAsnSerAspLeu..... 990
3499 GCAAGGCGTAGTCACCAACTTCGGTTCGGGATTCCAGATTTCATCGCG 3548
991 .....AsnProGlyThrThrGlnMetLysAspAsnPheAspIleGly 1004
||| |||:||||:||||:|||||
3549 TGCTGGCAGCGCGGGGACCA.....GGA 3574
1005 GluValCysProTrpGluValTyrAspLeuThrProGlyProValProSe 1021
||| |||:||||:||||:|||||
3575 AAC.....AGCCTCGCTCTCTGTACCGCGCGCGCTCCGCC 3612
1021 rGlu.....SerLysValGlnLysHisValSerIleValAlaSerGlu 1036
||| |||:||||:||||:|||||
3613 GCAACACCTGCAGATGTCGCCCTGCACCTGAGCACCTTCACGAGGAGT 3662
1036 etGluLysAsnProThrPheSerLeuLysGluLysSerHisHisLysPro 1052
||| |||:||||:||||:|||||
3663 CCATCTCCCTCCTGGGAGGACATCGATGATGACAGT..... 3700
1053 LysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAl 1069
||| |||:||||:||||:|||||
3701 .....GAGAGATTCAAGCTCCT 3717
1069 agluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
||||:||||:||||:|||||
3718 GCAGGAGTTCGTGTACGAGCGCGAAGGGAACACCGAAGAGATGAATTGG 3767

```


1502 TTTCCAGTGTGGCTACCTGGACGACCTCTTTGGAAAAACCCCAATTTAAGAA 1551
297ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
1552 AGTGTGCACAGNAATGAAGCTTGAAGAAAAACTATGTCCAGGACAGCA 1601
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
1602 AAATGGGATTGTGCATCAATGCCATCTATGCCATGSCACATGGGCTGCAG 1651
328 sLysCys...His.....LeuAsnAsnSerGluCysMe 338
1652 ACATGCACCATGCTCTGTGTCCCGGCCATGTGGGCGCTGTGTGATGCTAT 1701
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
1702 GAACCCATGTGGCAGGAACTCTGTGATTTCCTCATCAATCCTCTT 1751
347 347
1752 TTTCTCGAGTGTCTGGAGAGGAGGTGTGTTCATGAGAAGGGGATGCT 1801
348 ...GlyAlaTyrGluCystle..... 353
1802 CCGGAAGGTATGACATTATGAATCTGCAGTACACAGAAAGCTAATCGCTA 1851
354CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
1852 TGACTATGTCCAGTGGGACCTGGCATGAGGAGTGTGTAATATTGATG 1901
368 snPheArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
1902 AITACAAAATCCAG.....ATGACAAAAGCGAATGCTACGATCT 1942
385 ValSerGluGlu..... 388
1943 GTGTGCAGTGTAGCCCTTGCTTAAGAGGTGAGATTAAAGGTACATACGGAAGG 1992
389AlaTyrValCysLeuProCysArgGluGly..... 398
1993 AGAAGTGTAGCTGCTGCTGGATCTGCACGGCTGCANAGAGAAATGAGTTG 2042
399CysProPheCys..... 402
2043 TGCAGGACGAGTTACCTGTCAGAGCCTGTGACCTGGGTGTGTGGCCCAAC 2092
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2093 GCAGAGCTACAGAGGTGTGAGCCCAATCTCTCGCTTATCTTGATGGAG 2142
418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
2143 TGACATAGAAATCTATAGCCATCGCCTTTCTTGCCCTGGGCATCTCG 2192
430 euAspPheValSerMetLeuValTyrHisPheArgLysAlaLysSer 446
2193 TGACGCTGTTGTACCCCTCATCTTCGTCTCTACCGGACACACCCCGTG 2242
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLe 463
2243 GTCAATCTCCAGTAGGAGCTCTGTATATATATATCTGTGGCTGTATTTT 2292
463 uLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
2293 CTTGGCTATGTGGCCCTTTACCCCTCATCGCCAAACCTACTACCAAT 2342
480 rGcysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
2343 CTTGCTACCTCAGCGCCTCTCTAGTTGGCCTCTCTTCTGCCATGTGCTAC 2392
497 GlyThrValThrLeuLysHisArgValLeuLysVal..... 509
..... 3154

IMMEDIATE SOURCE:
 CLONE: 45-A
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 377..3973
 OTHER INFORMATION:
 US-08-455-602-1

alignment_scores:
 Quality: 232.00 Length: 1234
 Ratio: 0.480 Gaps: 51
 Percent Similarity: 39.141 Percent Identity: 19.044

alignment_block:
 US-09-775-181-2 x US-08-455-602-1 ..

Align seg 1/1 to: US-08-455-602-1 from: 1 to: 4300

```

153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
|||||
853 GATCGGCCCTGGCTCGCTGTCAGTCCAGTTCAGTCCAGAAATCTTCCTCC 902
169 OsErIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
|
903 A.....GCTGTTCGACATCCACACAGATCGCCTATTCTGCC 937
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
|||
938 ACAAGCATAGACTGAGTGACAAAACCTTTGTACAAATCTTCTGAGGGT 987
198 .....
988 GGTCCCTCTGACACTTTGCAGCGCAAGGCGGATGCTCGACATAGTCAAGC 1037
201 eLeuLeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrL 218
|||||
1038 GTTACAAGTGGACCTATGCTCAGCAGTCCACACAGAGGGAA..... 1080
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
1081 .....TTACGGCGAGAGTGGAAATGGATGCTTT 1107
231 .....ArgProHisLeuHisArgArg...GlyPr 239
|||||
1108 CAAAGAACTGGCTGCCAGGAAGGCT...CTGCATCGCACACTCGGACA 1154
239 oAsn..... 240
1155 AAATCTACAGCAATGCTGGCGAGAAGAGCTTTGACCGGCTCCTGCGTAAA 1204
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
|||||
1205 CTCGGGAGCGCTTCCCAAGC...CAGGGTTGTGGTCTGCTCTCGGA 1251
252 ArgLysAspGlyLeuGlyGlyAspLysSerHis...PheLysTrp..... 265
|||||
1252 GGGCATGACAGTGGCGGCTTACTCAGTCCCATGCGCGCGCTGGCGCTCG 1301
266 .....SerProTrpTyrLeuGluCysGluAsnGlySer..... 276
|||||
1302 TGGCGAGTTCTCACTCATTTGGAAGTGTATGGATGGCGACAGAGATGAA 1351
277 .....TyrLysPro..... 279
1352 GTCATCGAAGGCTATGAGTGGAGCCCAAGGAGGATCACAATAAAGCT 1401
280 .....Gly 280
1402 TCAGTCTCCAGAGGTGAGGTCAATTTGATCACTTCTCTGAAGCTGAGGC 1451
281 TrpLeuValThr.....LeuSe 286
|||

```

```

1452 TGGACACCAACACAAAGGAATCCTTGTTCCCTGAGTTCTTGCAACATCGC 1501
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu.... 296
|||||
1502 TTCCAGTGTGCGCTACCTGGACACCTCTTGGAAACCCCAACTTTAAGAA 1551
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
|||
1552 AGTGTGCACAGGAATGAAAGCTTGGAAAGAAACTATGTCCAGGACAGCA 1601
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
|||||
1602 AAATGGGATTTGTCAATCAATGCCATCTATGCCATGGCATGGGCTGCAG 1651
328 sLysCys..His.....LeuAsnAsnSerGluCysMe 338
|||
1652 AACATGCACCATGCTCTGTGTCGCCGCCATGTGGCCTGTGTGATGCTAT 1701
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
|||
1702 GAAACCCATTTGATGCGAGGAGCTCTCTGGATTCTCTCAATCAATCTCTT 1751
347 ..... 347
1752 TTGTCGGAGTGTCTGGAGAGGAGTGTGTTTCGATGAGAAGGGGATGCT 1801
348 ...GlyAlaTyrGluCysIle..... 353
1802 CCCGGAAGGTATGACATTTATGAATCTGCACAGACAGAGTAATCGCTA 1851
354 .....CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
|||||
1852 TGACTATCTCCACGTGGGACCTGGCAGTGAAGGAGTCTGAATATTGATG 1901
368 snPheArgArgGlyProaspGlnHisIleSerGlySerThrLysasp 384
|||||
1902 ATTACAAATCCAG.....ATGAACAAAGCGGAATGGTACGATCT 1942
385 ValSerGluGlu..... 388
1943 GTGTCCAGTGGACCTTGTAAAGGTCAGATTAAAGTTCATACGGAAGG 1992
389 .....AlaTyrValCysLeuProCysArgGly..... 398
|||||
1953 AGAAGTGAGTGTCTGCTGATCTGCACGCGCTGCAAGAGAATGAGTTTG 2042
399 .....CysProPheCys..... 402
2043 TGCAGGACGAGTTCCACCTGCAGAGCTGTGACCTGGGTGTGGCCCAAC 2092
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
|||||
2093 GCAGAGCTCAGAGCTGTGACCCCATTCCTGTCGTTATCTTGAGTGGAG 2142
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
|||||
2143 TGACATACAAATCTATCATAGCCATCGCCTTTTCTGCTGGCATCCCTCG 2192
430 euAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSer 446
|||||
2193 TGACGCTGTTTGCACCTCATCTTCGTTCTGTTACCGGGACACACCCG 2242
447 IleArgAlaSerGlyLeuLeuGluThrIleLeuPheGlySerLe 463
|||||
2243 GTCAATCTCCAGTAGGAGCTGTGCTATATCATCTTCTGGCTGGTATTTT 2292
463 uLeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheA 480
|||||
2293 CCTCGGTATGTGTGCCCTTTTCACTCCCTCATCGCAAACTACTACACAT 2342
480 rgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
|||||
2343 CTTGCTACCTCCAGCGCTCTCTAGTTGGCTCTCTCTGCGCATGTGCTAC 2392

```



```
3701 .....GAGAGATTCAAGCTCCT 3717
1069 aGluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
3718 GCAGGAGTTCGTGTACGCGGAGGGAACCCGAGGAGATGATTTGG 3767
1083 luAspGluLysLeuIleSerLysThrProValLeuProGluAfrAla 1099
3768 AGAGGAGGAGGACCTGCGCACAGCCAGCAAGCTGACCCCTGAGGATTCT 3817
1100 LysGluLysAsnGlyClyGlnProArgAlaAlaAsnValCysAlaGlyG 1116
3818 CCGGCCCTACGCCCTCTTCTCTCCGAGATTCCGTGCGCTTGGCAG 3867
1116 nSerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluA 1131
3868 CTCAGTCCCGCAGTTCCTCCGATCTGAGTCGGTCTCTCCACCCCTCCAA 3917
1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSer 1147
3918 ATGTAACCTACGCTCTGTCTCATCTGAGGAGCTACAAAGCAAGCTCTTCC 3967
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-157-1

```
seq_documentation_block:
; Sequence 1, Application US/08465157
; Patent No. 5869609
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE
; TITLE OF INVENTION: RECEPTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/465,157
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/041,538
; FILING DATE:
; APPLICATION NUMBER: US/07/648,481
; FILING DATE:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 206-623-6793
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
; ORIGINAL SOURCE:
; ORGANISM: Rattus norvegicus
; TISSUE TYPE: Cerebellum
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; OTHER INFORMATION:
US-08-465-157-1

alignment_scores:
Quality: 232.00 Length: 1234
Ratio: 0.480 Gaps: 51
Percent Similarity: 39.141 Percent Identity: 19.044

alignment_block:
US-09-775-181-2 x US-08-465-157-1 ..
Align seg 1/1 to: US-08-465-157-1 from: 1 to: 4300

153 AspLeuAspTrpTrpTyrGlnAlaLeuValTrpSerLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 GATCGGCCCTGGCTCCAGCTCTGTGGCCATTCAAGTCCAGAAATCTCTCC 902
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
903 A.....GCTGTTCGACATCCACAGATCGCTTATTCTTCTGCC 937
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 ACAAGCATAGACCTGAGTCACAAACTTTGTACAATACTTCTCTGAGGT 987
198 .....GluSerArgIle 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
988 GGTCCCTTCTGACACTTTTCAGGAGGCGGATGCTCGACATAGTCAAGC 1037
201 eLeuLeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1038 GTTAACTGGACCTATGTCTCAGCGTCCACAGAGGGAA..... 1080
218 euGluThrGluTrpPheHisGlyLeuArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 .....TTACGGGAGAGTGGAAATGATGCTTT 1107
231 .....ArgProHisLeuHisArgArg....GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 CAAGAACTGGCTGCCAGGAGGCT...CTGCATCGCACACTCGGACA 1154
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 AAATCTACAGCAATGCTGGCGAGAGAGCTTTGACCGGCTCTGCGTAA 1204
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1205 CTCGGGAGGGCTTCCCAAGGC...CAGGGTTGGGTCTGCTCTCTCGGA 1251
252 ArgLysAspGlyLeuGlyAspLysSerHis..PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1252 GGGCATGACAGTGGCGGCTTACTAGTCCCATGCGCGCCCTGGCGCTCG 1301
256 .....SerProTrpLeuGluCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1302 TGGCGAGTTCTCACTCATTTGGAGTGATGGGCGACAGAGATGAA 1351
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1352 GTCATCGAAGGCTATGAGTGGAGGCCAACCGAGGAGGATCACAATAAGCT 1401
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1402 TCAGTCTCCAGAGTCAAGTCACTTTTGATGACTACTTCTGAAGCTGAGGC 1451
```



```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: 45-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 377..3973
; PCT-US91-09422-1
```

alignment_scores:

```
Quality: 232.00 Length: 1234
Ratio: 0.480 Gaps: 51
Percent Similarity: 39.141 Percent Identity: 19.044
```

alignment_block:

```
US-09-775-181-2 x PCT-US91-09422-1 ..
```

```
Align seg 1/1 to: PCT-US91-09422-1 from: 1 to: 4300
```

```
153 AspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPr 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 GATCGGCCCTGCTCCAGCTCTGCGCCATTCAAGTCCAGAACTTCTCC 902
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 oSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu..... 183
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
903 A.....GCTGTCGACATCCACAGATCGCCTATTCTGCC 937
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....SerAlaProAlaProGlnValPheLeuGlnAlaThrArgGlu... 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
938 ACAAGCATGACCTGAGTGACAAAACTTTGTACAAATACATTCTCTGAGGGT 987
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 .....GCTGTCGACATCCACAGATCGCCTATTCTGCC 937
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
988 GGTCCCTTCGACATTTGCGAGCAAGGCGATGCTCGACATAGTCAAGC 1037
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eLeuLeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrL 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1038 GTTACAACTGGACCTATGCTCAGCAGTCCACAGAAAGGAA..... 1080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 euGluThrGluTrpPheHisGlyLeuArgArgLysTrp..... 230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1081 .....TTACGGCGAGAGTGGAAATGATGCTTT 1107
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....ArgProHisLeuHisArgArg...GlyPr 239
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 CAAAGAACTGGCTGCCAGGAAGGCT...CTGCATCGCACACTCGGACA 1154
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 oAsn..... 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 AAATCTACAGCAATGTGGCGAGAAGAGCTTTGACCGGCTCTCTCGGTAA 1204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 .....GlnGlyProArgGlyLeuGlyHisSerTrpArg 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1205 CTCGGGAGCGGCTTCCCAAGC...CAGGGTTGTGCTGCTCTCGGA 1251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ArgLysAspGlyLeuGlyAspLysSerHis...PheLysTrp..... 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1252 GGCATGCACAGTCCGGGCTTACTGAGTCCATGCGCCGCTGGCGTCG 1301
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 .....SerProTrpTyrLeuGlyCysGluAsnGlySer..... 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1302 TGGGCGAGTTCCTCACTTGGAGGTGATGGATGGCGAGACAGATGAA 1351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....TyrLysPro..... 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1352 GTCATCGAAGGCTATGAGGTGGAAAGCAAGGAGGAGGATCAATAAAGCT 1401
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 .....Gly 280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1402 TCAGTCTCCAGAGGTGAGGTCACTTTGATGACTACTTCTCTGCAAGCTGAGGC 1451
```

```
281 TrpLeuValThr.....LeuSe 286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1452 TGGACACCAACACAGGAATCTTGTTCCTTCCTGAGTTCTGGCAACATCGC 1501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 rSerAlaIleTyrGlyLeuGln.....ProAsnLeu..... 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1502 TTCCAGTTCGCTACCTACCTGGACACCTCTTGGAAACCCCAACTTTAAGAA 1551
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 .....ValProGluPheArgGlyValMetLysValAspIleAsnLeuGln 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 AGTGTGCACAGGAATGAAGCTTGAAGAAACTATGTCCAGGACAGCA 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 LysValAspIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHi 328
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 AAATGGGATTTGCATCAATGCCATGCTATGCCATGGCAGATGGCTGCAG 1651
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 sLysCys...His.....LeuAsnAsnSerGluCysMe 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 AACATGCACCATCTCTGTCTCCGGCCATGTGGGCTGTGTGATGCTAT 1701
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 t...ProIleLysGly.....LeuGlyPheValLeu..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1702 GAAACCCATTGATGGCAGGAAGCTCTGGATTCTCCTCATCAATCCTCTT 1751
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 ..... 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1752 TTTCGAGGTCTGAGAGAGGAGTGTGTTGATGAGAGGGGGATGCT 1801
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 ...GlyAlaTyrGluCysIle..... 353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1802 CCCGGAAGGTATGACATTTATGAATCTGCAGTACACAGAAGCTAATCGCTA 1851
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 .....CysLysAlaGlyPheTyrHisProGlyValLeuProValAsnA 368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1852 TGACTATGTCACGTGGGACCTGGCATGAAGAGTGTCTGAATATTGATG 1901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 snPheArgArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1902 ATTACAAATCCAG.....ATGAACAAAGCGGAATGGTACGATCT 1942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 ValSerGluGlu..... 388
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1943 GTGTGAGTGAGGCTTGTCTTAAAGGTCAGATTAAAGTCTACGGAAGG 1992
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 .....AlaTyrValCysLeuProCysArgGly..... 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1993 AGAAGTGAGTGTCTGTGGATCTGCAGCGCTGCAAGAGAATGAGTTTG 2042
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 .....CysProPheCys..... 402
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2043 TGCAGGACGAGTTTCACTGCAGAGCCTGTGACCTGGGGTGGTGGCCAAC 2092
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 AlaAspAspSerProCysPheValGlnGluAspLysTyrLeuArg..... 417
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2093 GCAGAGTCAAGGCTGAGCCCATCTCTGTCCTGCTATCTTGCAGTGGAG 2142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuL 430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2143 TGACATAGATCTATCATAGCCATCGCTTTCTGCTGGGCATCTCTCG 2192
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 euAspPheValSerMetLeuValTyrHisPheArgLysAlaLysSer 446
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2193 TGACGCTGTTGTCACCTCATCTGCTTGTACCGGGACACACCGGTG 2242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
447 IleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLe 463
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2243 GTCAATCTCCAGTAGGAGCTCTGCTATATCATCTTGGCTGGTATTT 2292
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 uLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheA 480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2293 CCTCGGCTATGTGCGCTTTCACCTCTACCGCAACACTACTACCACAT 2342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 rgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyr 496
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
1053 LysAlaAlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAl 1069
      ::|||:::
3701 .....GAGAGATTCAAGCTCCT 3717
      ::|||:::
1069 aGluValCysLeuTrpGluSerGlnGly.....GlnSerIleLeuG 1083
      ::|||:::
3718 GCAGAGTGTGTACAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3767
      ::|||:::
1083 luAspGluLysLeuLeuSerLysThrProValLeuProGluArgAla 1099
      ::|||:::
3768 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3817
      ::|||:::
1100 LysGluGluAsnGlyGlnProArgAlaAlaAsnValCysAlaGlyG1 1116
      ::|||:::
3818 CCTGCCCTGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3867
      ::|||:::
1116 nSerGluGluLeuProPro.....LysAlaValAlaSerLysThrGluA 1131
      ::|||:::
3868 CTCAGTCCCGAGTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3917
      ::|||:::
1131 snGluAsnLeuAsnGlnIleGlyHisGlnGluLysLysThrSerSerSer 1147
      ::|||:::
3918 ATGTAACCTACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3967
      ::|||:::
```

seq_name: /cqn2_6/ptodata/2/ina/6B_COMB.seq:us-08-660-148-1

```
seq_documentation_block:
; Sequence 1, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,148
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-9419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
```

```
LOCATION: 460...3999
US-08-660-148-1

alignment_scores:
Quality: 216.50 Length: 1067
Ratio: 0.454 Gaps: 42
Percent Similarity: 44.705 Percent Identity: 18.650

alignment_block:
US-09-775-181-2 x US-08-660-148-1 ..
Align-seg 1/1 to: US-08-660-148-1 from: 1 to: 4207

219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
      ::|||:::
1456 GATGTCAGTGGTTGTATGATATTATCTGAAGCTCCGGCCAGAAACAAA 1505
      ::|||:::
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg. 251
      ::|||:::
1506 CCACCGAAACCT.....TGGTTTCAAGAATTTTGGCAGC 1540
      ::|||:::
252 .....ArgLysAspGlyLeuGlyAspLysSerHisPhe 263
      ::|||:::
1541 ATCGTTTTCAGTCCCGACTGGAAGGGTTTCCACAGGAGAACAAATAC 1590
      ::|||:::
264 LysTrpSerProTyrLeuGluCysGluAsnGly.....MetLysValAspI1 308
      ::|||:::
1591 AACAGACT.....TGCATAGTCTCTGACTCTGAAAC 1625
      ::|||:::
276 .....SerTyrLysProGlyTrpLeuVal...ThrLeuSerS 287
      ::|||:::
1626 ACATCATGTTCCAGGATCCAAATGGGATTGTGATCAACGCCATCTATT 1675
      ::|||:::
287 erAlaIleTyrGlyLeu.....GlnProAsnLeuValProGluPhe 300
      ::|||:::
1676 CGATGGCTATGGGCTCCACACATGCAGATGTCCCTCTGCCCGAGCTAT 1725
      ::|||:::
301 ArgGlyVal.....MetLysValAspI1 308
      ::|||:::
1726 GCAGGACTCTGTGATGCCATGAACCAATTTGATGGACGGAACNTTTTGA 1775
      ::|||:::
308 eAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPheS 325
      ::|||:::
1776 GTCCCTGATGAAACCAATTTTACTGGGGTTTCTGGAGATACGATCCTAT 1825
      ::|||:::
325 erGlyThrHisLysCysHisLeuAsnSerGluCysMetProIleLys 341
      ::|||:::
1826 TCGATGAGAATGGAGACTCTCCAGAAAGGATGAATAATATTCAG 1875
      ::|||:::
342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyPh 358
      ::|||:::
1876 GAAATGGGA.....AAAGATTACTT 1895
      ::|||:::
358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
      ::|||:::
1896 TGATTATATCAAGCTTGGAAAGTTGGGACAATGGAGAATTAATAATGGATG 1945
      ::|||:::
375 spGlnHisIle.....SerGlySerThrLysAspValSerGlu 387
      ::|||:::
1946 ATGATGAAGTATGTTCCAGAAAGCAACATCATCAGATCTGTGTGCAGT 1995
      ::|||:::
388 Glu..... 388
      ::|||:::
1996 GAACCATGTGAGAAAGCCAGATCAAGTGTCCGAAAGGAGNAGTCAG 2045
      ::|||:::
389 .....AlaTyrValCysLeuProCysArgGlyGly..... 398
      ::|||:::
2046 CTGTTGTTGGACCTGTACACCTTGTAGGAGAGATGAGTATGCTTTTGATG 2095
      ::|||:::
399 .....CysProPheCys.....AlaAspAsp... 405
      ::|||:::
2096 AGTACACATCAAGGCATGCCAACTGGGGCTTTGGCCCACTGATGATCTC 2145
      ::|||:::
```

406 SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
::: ||| :::: |||||
2146 ACAGGTTGTACCTGATCCAGTACAGTATCTTCGATGGGTGACCCCTCA 2195
418LeuAlaIleSerPheGlnGlyLeuCysMetLeuLeuAspPheV 433
|||::: ||| :::: |||||
2196 ACCATTGGCAGTGGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2245
433 aIserMetLeuValValTyrHisPheArgLysAlaLysSerIleArgAla 449
|||::: ||| :::: |||||
2246 TTGTTACTAGTCTTCATCATTTACCGGTATACACCATAGTCAAGTCC 2295
450 SerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTy 466
|||::: ||| :::: |||||
2296 TCAAGCAGGGAATCTGCTACATTTATCTTGTGCTGCTGCTGCTGCTGCT 2345
466 rPheProValIleLeuTyrPheCysProSerThrPheArgCysIleL 483
|||::: ||| :::: |||||
2346 CTATGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2395
483 euLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrVal 499
|||::: ||| :::: |||||
2396 TTCAGAGAAATTGGCATTTGCTCTCTCCCAAGCCATGAGTACTACGCTT 2445
500 ThrLeuLysLeuHisArgValLeuLysVal..... 509
::: ||| :::: |||||
2446 GTAACAAGAACCAACGATTTCAAGATCTTGGCTGGCAGCAAGAAGAA 2495
510PheLeuSerArgThrAlaGlnArgIleP 519
2496 GATCTGTACCAAAAGCCAGATTTATGATGCTGCTGCTGCTGCTGCTGCT 2538
519 roTyrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeu 535
2539CTA 2541
536 ValValPheTrpPheLeuIleGlyTrpThrSerValCysGlnAsnLe 552
|||::: ||| :::: |||||
2542 GTGATTTGCTTTCATCTCATA.....TGCATCCAGTT 2573
552 uGluLysGlnIleSerLeu.....IleGlyGlnG 562
2574 GGGCATCATGTTGCTGCTCTTTATGAGGCTCTCTGACATTAATGTCATG 2623
562 lYlYsThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
|||::: ||| :::: |||||
2624 ACTACCAAGCATTCGAGAAGTCTACCTGATCTGTAACACCAACCACTA 2673
579 AspTyrMetThrAlaValAla...GluPheLeuPheLeuLeuTrpGlyVa 594
|||::: ||| :::: |||||
2674 GGAGTTGTACCTCCACTTGGATACAAATGGATTGTTGATTTGAGTGCAC 2723
594 lTyrLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProA 611
|||::: ||| :::: |||||
2724 CTTCATGCTTCAAGACCAAGAAATGTTCCAGCTAACTTCAACGAGCCA 2773
611 rGlyMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePhe 627
|||::: ||| :::: |||||
2774 AGTATATGCTTTCATATGATACAGACCTGCATTTATATGCTAGCTTTT 2823
628 HisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetLe 644
|||::: ||| :::: |||||
2824 GTGCCAATCTACTTT.....GGCAGCAACTACAAAAT 2855
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
|||::: ||| :::: |||||
2856 CATCACCATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2905
661 euLeuLeuIleProLys.....PheSerHisSerSerAsnAsn 673
|||::: ||| :::: |||||
2906 GCATGTTTGTCCGAAGTGTATCATCATCTCTGCTGCTGCTGCTGCTGCTGCT 2955

674 ProArgAspAspIleAlaThrGlu...AlaTyrGluAspGluLeuAspMe 689
|||::: ||| :::: |||||
2956 GTGCGCAGCGCTTACCACATCTACCTGGTGGCATGTAGGGGA 3005
689 tGlyArgSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlaT 704
|||::: ||| :::: |||||
3006 TGGCAAGTATCTCTCGCAGCAGATCCAGCAGCTTAGTTCGAATGGAATCC 3055
704 rPserGluHisSerLeuAspProGluAspIleArgAspGluLeuLeuLys 720
|||::: ||| :::: |||||
3056 GGAAGAAGAGGGCTCTCTGGGAAACCTTAAGTTCCAAATGGAATCC 3105
721LeuTyrAl 723
|||::: ||| :::: |||||
3106 GTCAGTGGGCCAGAAATGAGAAGAGCAGCGGGGCGACACCTGTGGCA 3155
723 agLnuGluIleTyrLysArgLysLysMetIleThrAsnAsnProHisL 740
|||::: ||| :::: |||||
3156 GCGCTGTCCATCCACATCAACAGAAAGAAACCCCAACCAACGCGCG 3205
740 euGlnLys.....LysArgCysSerLysLysLysLysLysLysLys 753
|||::: ||| :::: |||||
3206 TCATCAAGCCCTTCCCAAGACGAGAGCGCTGGCTGGCGCTGGC 3255
754 IleMetArgArgIleThrGluIleProGluThrValSerArgGlnCysSe 770
3255 3255
770 rLysGluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaL 787
|||::: ||| :::: |||||
3256GCTGGCGCAGGC...GGGAGCGCTGGGGCGTGGGGG 3289
787 euIleArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLys 803
|||::: ||| :::: |||||
3290 CCACGGCGGTGCGGGCTGCGCAGCGCGCGCCAGCGCGCGCGCGAGTCC 3339
804 GluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerTh 820
|||::: ||| :::: |||||
3340 CCAGACGCGCGCCCAAGCGCTGAT..... 3366
820 rTyrAspHisValArgAspGlnThrGluLysSerSerSerLeuProThrG 837
|||::: ||| :::: |||||
3367GATGGCGGAGGCTGAGGAGCACTTCCCGGCGC 3400
837 luserGlnGluGluThrThrGluAsnSerThrLeuGluSerLeuSer 853
|||::: ||| :::: |||||
3401 CCGCGCGCGCGCTCACCCTCGCCCATCAGACGCTCAGCCACCGCGCG 3450
854 GlyLysLysLeuThrGlnLysLysGluAspSerGluAlaGluSerTh 870
|||::: ||| :::: |||||
3451 GGC.....TCGGCCAGCGCGCAGCAGCAGATGTGCCGCTGCT 3488
870 rGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLeu...Se 886
|||::: ||| :::: |||||
3489 GCATCGGAGGCTGCGCGCGCAGCAGCTCTCGCAGGCGCTCCCTCATGG 3538
886 rSerGluLysLysThrGlyHisProArg...ThrSerMetLeuGlnLysSer 902
|||::: ||| :::: |||||
3539 AGCAGATCAGCAGTGTGTCACCGCTTCACGGCCACATCAGCAGGCTC 3588
903 LeuSerValIleAlaSerAlaLysLysThrLeuGlyLeuAlaGlyL 919
|||::: ||| :::: |||||
3589 AACTCCATGATGCTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCG 3638
919 yThrGlnThrAlaGlyValGluArgThrLysSerGln..... 932
|||::: ||| :::: |||||
3639 GCTCTGCTGCTCTACCTGATCCCAAGAGATCCAGTGTGCCACGACCA 3688
933 ...LysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSe 948
|||::: ||| :::: |||||
3689 TGACGACCTTTCGGAATCCAGCTCTGCGCGCGCATCGAAGTCACGGCG 3738
948 rAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAlaG 965

[illegible]


```

; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3912
; OTHER INFORMATION: /product= "HUMAN MGLUR5A"
; US-08-367-264-7

alignment_scores:
  Quality: 215.00      Length: 1084
  Ratio: 0.461        Gaps: 42
  Percent Similarity: 42.989  Percent Identity: 18.358

alignment_block:
US-09-775-181-2 x US-08-367-264-7 ..
Align seg 1/1 to: US-08-367-264-7 from: 1 to: 4085

219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
      ::::::::::::::::::::||| |||||:::
1366 GATGTCACAGTGGTTGATGATATATCTGAAGCTCCGCCAGAACAA 1415
      ::::::::::::::::::::||| |||||:::

235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg 252
      ::::::::::::::::::::||| |||||:::
1416 CCACCGAAACCCCT.....TGGTTTCAAGAAATTTGGCAGC 1450

252 rLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerProPro 268
      ::::::::::::::::::::||| |||||:::
1451 ATCGT.....TTTCAGTGGCCGACTGGAA 1473

269 TyrLeuGlnCysGluAsnGlySerTyr..... 277
      ::::::::::::::::::::||| |||||:::
1474 GCGTTTCCACAGAGACAGCAACAAATACAAAGACTTGCATAGTCTCT 1523

278 .....LysProGlyTrpLeuVal. 283
      ::::::::::::::::::::||| |||||:::
1524 GACTCTGAAACACATCATGTTCAGGATTCCAAATGGGATTTGTGATCA 1573

284 .....ThrLeuSerSerAlaIleTyrGlyLeu.....GlnProAsnLeu 296
      ::::::::::::::::::::||| |||||:::
1574 AGCCATCTATTTCGATGGCTATGGCTCCACAAATGCAGATGTCCCTTC 1623

297 ValProGluPheArgGlyVal.....Me 304
      ::::::::::::::::::::||| |||||:::
1624 TGCCCGAGGTATGCAGGACTCTGTGATGCCATGAAGCCAAATGTATGGAGC 1673

304 tLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSera 321
      ::::::::::::::::::::||| |||||:::
1674 GAACTTTTGGAGTCCCTGATGAAACCAATTTTACTGGGGTTCTGGAG 1723

321 spGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCys 337
      ::::::::::::::::::::||| |||||:::
1724 ATACGATCTATTTCGATGAGATGAGACTCTCCAGGAGGATGAAATA 1773

338 MetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCy 354
      ::::::::::::::::::::||| |||||:::
1774 ATGAATTTCAAGAAATGGGA..... 1794

354 sLysAlaGlyPheTyrHisProGlyValLeuProValAsnAsnPheArg 371
      ::::::::::::::::::::||| |||||:::
1795 .AAGATTACTTTGATATATCAACGTTTGAAGTTGGACATGGAGAT 1843

371 rArgGlyProAspGlnHisIle.....SerGlySerThrLys 383
      ::::::::::::::::::::||| |||||:::
1844 TAAATATGATGATGATGAAGTATGGTCCAGAAAGCAACATCATCAGA 1893

```

```

384 AspValSerGluGlu..... 388
      ::::::::::::::::::::||| |||||:::
1894 TCTGTGTGCAGTGAACCATGTGAGAAAGCCAGATCAAGGTGATCCGANA 1943

389 .....AlaTyrValCysLeuProCysArgGluGly.... 398
      ::::::::::::::::::::||| |||||:::
1944 GGGAGAAGTCAGCTGTGTGGACCTGTACACCTTGTAGGAGAATGAGT 1993

399 .....CysProPheCys..... 402
      ::::::::::::::::::::||| |||||:::
1994 ATGCTTTTGTGATGATACATCAAGGCATGCCAACTGGGGTCTTGGGCC 2043

403 AlaAspAsp....SerProCysPheValGlnGluAspLysTyrLeuArg... 417
      ::::::::::::::::::::||| |||||:::
2044 ACTGATGATCTCACAGGTGTGACTTGATCCAGTACAGATATCTTCGATG 2093

418 .....LeuAlaIleSerPheGlnGlyLeuCysMetL 429
      ::::::::::::::::::::||| |||||:::
2094 GGGTGACCTGAACCAATGTCAGCTGTGTGTTCCTGCCTGGCTGCC 2143

429 euLeuAspPheValSerMetLeuValValTyrHisPheArgLysAlaLys 445
      ::::::::::::::::::::||| |||||:::
2144 TGGCCACCTGTTTGTGTACTGTAGTCTTCATCATTTACCGTGATACACA 2193

446 SerIleArgAlaSerGlyLeuIleLeuGluThrIleLeuPheGlySe 462
      ::::::::::::::::::::||| |||||:::
2194 GTAGTCAGTCTCAAGCAGGGAACCTCTGTACATATCTTGTCTGGCAT 2243

462 rLeuLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrP 479
      ::::::::::::::::::::||| |||||:::
2244 CTGCTGGGCTACTTATGTATCTTGTCTCATTTGCGAAGCCCAACAGA 2293

479 heArgCysIleLeuLeuArgTrpAlaArgLeuGlyPheAlaThrVal 495
      ::::::::::::::::::::||| |||||:::
2294 TTTACTGTACTCTCAGAAATGGCATTTGGTCTCTCCCGACCATGAGC 2343

496 TyrGlyThrValThrLeuHisArgValLeuLysVal..... 509
      ::::::::::::::::::::||| |||||:::
2344 TACTCAGCCCTTGTAACAAGACCAACCGTATTCAGAGATCCTGGCTGG 2393

510 .....PheLeuSerArgThr 515
      ::::::::::::::::::::||| |||||:::
2394 CAGCAAGAAGAAGATCTGTACCCCAAGCCAGATTCATGATGCTGTG 2443

515 LglnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAla 531
      ::::::::::::::::::::||| |||||:::
2444 CCCAG..... 2448

532 ValIleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVa 548
      ::::::::::::::::::::||| |||||:::
2449 .....CTAGTGATTGCTTTCATCTCATA..... 2472

548 lCysGlnAsnLeuGluLysGlnIleSerLeu..... 558
      ::::::::::::::::::::||| |||||:::
2473 .TGATCCAGTTGGGCATCATCGTTGCCCTCTTTATAATGGAGCCTCTG 2521

559 .....IleGlyGlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeu 574
      ::::::::::::::::::::||| |||||:::
2522 ACATAATGCATGACTACCAAGCATTCGAGAGTCTACCTGATCTGTAAAC 2571

575 lLeAspArgTrpAspTyrMetThrAlaValAla...GluPheLeuPheLe 590
      ::::::::::::::::::::||| |||||:::
2572 ACCACCAACCTAGGAGTGTCTACCTCCACTTGGAAACAAATGGATTGTGAT 2621

590 uLeuTrpGlyValTyrLeuCysTyrAlaValArgThrValProSerAlaP 607
      ::::::::::::::::::::||| |||||:::
2622 TTTGAGCTGCACCTTCTATGCTTCAAGACCAAGAAATGTTCCAGCTAACT 2671

607 heHisGluProArgTyrMetAlaValAlaHisAsnGluLeuIleIle 623
      ::::::::::::::::::::||| |||||:::
2672 TCCCGGAGCCCAAGTATATCCCTTCACATGTACACGACCTGCATTATA 2721

```


2496 GATCTGTACCAAAAGCCAGATTCATGCTGCTGTGCCAG..... 2538
519 roTyrMetThrGlyArgValMetArgMetLeuAlaValIleLeuLeu 535
2539CTA 2541
536 ValValPheThrPheLeuIleGlyTyrThrSerValCysGlnAsnLe 552
2542 GTGATTGCTTCATTCATA.....TGCAATCCAGTT 2573
552 uGluLysGlnIleSerLeu.....IleGlyGlnG 562
2574 GGGCATCATGCTGCCCTCTTTATATGAGAGCTCTGACATAATGCATG 2623
562 lYsThrSerAspHisLeuPheAsnMetCysLeuIleAspArgTyr 578
2624 ACTACCAAGATTCGAGAATCTACCTGATCTGTAAACACCAACCTA 2673
579 AspTyrMetThrAlaValAla...GluPheLeuPheLeuTyrGlyVa 594
2674 GGAGTCTCACTCCACTTGGATACAAATGGATGTTGATTTGAGCTGCAC 2723
594 lTyrLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProA 611
2724 CTTCTATCGCTTCAAGACCAAGAAATGTTCCAGCTAACTTCAACGAGCCA 2773
611 rgTyrMetAlaValAlaValHisAsnGluLeuIleIleSerAlaIlePhe 627
2774 AGTATATCGCTTCCAAATGTACAGACCTGCATTTATATGCTAGCTTT 2823
628 HisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTyrMetLe 644
2824 GTGCCATCTACTTT.....GGCAGCAACTACAAAT 2855
644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
2856 CATCACCATGTGTTCTCGGTACGCTCAGTCCACAGTGGCCCTAGGCT 2905
661 euLeuLeuIleProLys.....PheSerHisSerSerAsnAsn 673
2906 GCATGTTGTGGCGAAGGTGTACATCATCTGCGCAACACAGAGAGAAC 2955
674 ProArgAspAlaLeuThrGlu...AlaTyrGluAspGluLeuAspMe 689
2956 GTGGCAGAGCCCTTACCACATCTACCGTGGTGGCATGTAGTGGGA 3005
689 tGlyArgSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlat 704
3006 TGGCAAGTCATCTCCGCGAGCCAGCAGATCCAGCAGCTAGTCAACCTGT 3055
704 rpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLys 720
3056 GGAAGAAAGGGGCTCTCTGGGAAACCTTAAGGTACAAAGACAGAGA 3105
721 Leu.....TyrAlaGlnLeuIleTyrLysArgLysLysMetIl 734
3106 CTGCCCCAGCAAGTCCGAAATAGAGTGTTCACCCCAAGGGAGTAT 3155
734 e.....ThrAsnAsnProHisLeuG 741
3156 GGGGAATGGTGGGAGACAACATGACAGTTCCCAATGGAAATCCCGTCA 3205
741 lnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgArg 757
3206 CGTGGGCCCAAGATGAGAAGACGAGCGGGGAGACACCTGTGGCAGCGC 3255
758 IleThr.....GluIlePro..... 762
3256 CTGTCCATCCACATCACAGAAAGAAACCCCAACCAACGCGCGTAT 3305
763GluThrValSerArgGlnCysSerLysGluAspL 774
3306 CAAGCCCTTCCCCAAGACGAGGAGACCGGTGGCCTGGCGCTGGCGCTG 3355

774 ysGluGlyAlaAspHisGly.....ThrAlaLysGlyThr 785
3356 GCGCAGCGGGAGCGTGGGGCGTGGGCCACGGCGGTGGCGCTGC 3405
786 AlaLeuIleArgLysAsnProGluSerSerGlyAsnThrGlyLysLe 802
3406 GCAGGCCCGCGCCAGCGGGCGCGAGTCCCGACAGCCCGCCCAAGGC 3455
802 rLysGluGluThrLeuLysAsnArgValPheSerLeuLysSerHis 819
3456 G..... 3456
819 erThrTyrAspHisValArgAspGlnThrGluGluSerSerLeuPro 835
3457CTGTATGATGTGGCGAGGCTGAGGAGCACCTTCCCG 3492
836 ThrGluSerGlnGluGluThrThrGluAsnSerThrLeuGluSerLe 852
3493 GCGCCCGCGCGCGCTCACCTCGCCCATCAGCACGCTGAGCCACCG 3542
852 uSerGlyLysLysLeuThrGlnLysLeuLysGluAspSerGluAlaGlu 869
3543 CGCGGGC.....TCGGCCAGCGCAGCGAGCATGTGCGCT 3580
869 erThrGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLeu 885
3581 CGGTGCACTCGGAGCTGTGGCGCGCAGAGCTCTCGCAGGGCTCCCTC 3630
886 .SerSerGluLysLysThrGlyHisProArg..ThrSerMetLeuGlnL 901
3631 ATGGAGCAGATCAGCAGTGTGGTACCCGCTTACGGCCCAACATCAGCA 3680
901 sSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeuAla 917
3681 GCTCAACTCCATGATGTGTCCACCGGGCCCGCAGCCCGCGTGGCG 3730
918 GlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGln..... 932
3731 CCGCGCTCTGCTCTACTGATGATCCCAAGAGATCCAGTGGCCACG 3780
933LysProLeuProLysAspLysGluThrAsnArgAsnHisSerA 947
3781 ACATGACAGACCTTGCAGAAATCCAGCCTTCCCGGCCCATCGAAGTCAC 3830
947 snSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnPro 963
3831 GGGCGCGCGCAGCCCGCGCGGCGCAGCGGCTGGGAGCGCGGCC 3880
964 AlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGlnAr 980
3881 GGGAGAGCCCGCGG.....CCGCTCCCGAG 3906
980 gValAsnProThrThrAlaAsnSerAspLeuAsnProGlyThrThrGlnM 997
3907 GCTGCGCGCG.....CCAGCCAGACC..... 3928
997 etLysAspAsnPheAspIleGlyLeuValCysProTyrGluValTyrAsp 1013
3929TCGAGGAGCTGTGG 3943
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValse 1030
3944 CTCTCACCCCGCGCTCCCTTCCAGACTCGGTGAGTCTGGGAGACACA 3993
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys..... 1045
3994 ACCCCCAACTCGCAGTGTGGAGTGGCGCTCTGTATCCGCTGTCTCC 4043
1046GluLysSerHisLysLysProLysAla 1054
4044 CAATATGACACTCTTATATAGAGATTTACACTCAGAGCTCTCTCGCGC 4093

1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVa 1071
:::||||
4094 TGTGA..... 4099
1071 lCysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLysLeuL 1088
||| |||:::|:::|
4100 .TGTCCCTGGAAAGCAGC..... 4117
1088 eulleSerLysThrProValLeuProGluArgAlaLysGluAsnGly 1104
||| |||
4118CGGCTGCGCGTGG..... 4132
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
|||::: ||
4133GAGCGGAGCCC 4143
1121 oProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleG 1138
||| |||:::|:::|
4144 CCG.....TGTTACACACACAAAT.....G 4166
1138 lYHisGlnGluLysLysThrSerSerGluGluAsnValArgGlySer 1154
|| :::::|:::|
4167 GCAAGCATAGCCGCTGTTACGCCGCCAGGGGAAAGTCCCAAGGCACC 4216

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-660-148-6

seq_documentation_block:
; Sequence 6, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/660,148
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REFERENCE/DOCKET NUMBER: X-9419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
US-08-660-148-6

alignment_scores:
Quality: 209.50 Length: 1085
Ratio: 0.436 Gaps: 43
Percent Similarity: 44.332 Percent Identity: 18.525
alignment_block:
US-09-775-181-2 x US-08-660-148-6 ..
Align seg 1/1 to: US-08-660-148-6 from: 1 to: 4303
219 GluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHi 235
||| |||:::|:::|
1456 GAUGUCAAGUGGUUGAUGAUUAUUGAAGCUCGCCGACGAGAAACAAA 1505
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArg. 251
:::|:::|:::|
1506 CCACGAAACCCU.....UGGUUCAAGAAUUUUGGCACC 1540
252ArgLysAspGlyLeuGlyGlyAspLysSerHisPhe 263
||| :::|:::|
1541 AUGGUUUUCAGUGCGCAGUGGAAGGUUCCACAGGAGAACAGCAAUAC 1590
264 LysTrpSerProTrpLysLeuGluCysGluAsnGly..... 275
:::|:::|:::|
1591 AACAAAGACU.....UGCAUAGUUCUGACUCUGAAAC 1625
276SerTyrLysProGlyTrpLeuVal...ThrLeuSers 287
||| |||:::|:::|
1626 ACAUCAUGUUCAGGAUCCAAUUGGGAUUUGAUCACGCGCAUCUAU 1675
287 erAlaIleTyrGlyLeu.....GlnProAsnLeuValProGluPhe 300
|| |||:::|
1676 CGAUGGCCUUAUGGGGUCCACAAUGCAGAUCCUCUGCCGCGCUAU 1725
301 ArgGlyVal.....MetLysValAspIle 308
|||:::|
1726 GCAGGACUCUGUGAUGCCAAUGAUGGAGCGGAAACUUUUGGA 1775
308 eAsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTrpPheS 325
:::|:::|:::|
1776 GUCCUGAUGAAACCAUUUACUGGGGUUUGGAGAUACGAUCCUAU 1825
325 erGlyThrHisLysCysHisLeuAsnSerGluCysMetProIleLys 341
:::|:::|
1826 UCGAUGAAGAUUGGAGACUCUCAGGAGGUAUGAAUAAUUAUUAAG 1875
342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyP 358
:::|:::|
1876 GAAUGGGA.....AAAGAUAUACUU 1895
358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
||| :::|:::|
1896 GAACAUGGAGAAAGCCACCAUGAUGGAGGAGAAUUAUUAUUAUGAUG 1945
375 spGlnHisIle.....SerGlySerThrLysAspValSerGlu 387
|||:::|
1946 AUGAUGAAGAUUGGUCCAAAGAAACCAUCAUCAUGAUGUGUGCAGU 1995
388 Glu..... 388
1996 GAACAUGGAGAAAGCCACCAUGAUGGAGGAGGAGAAUGGAGU 2045
389AlaTyrValCysLeuProCysArgGluGly..... 398
:::|:::|
2046 CUGUUGUUGGACCUACACCCUUGUAAGCAGAUCAUGUUCUUGAUG 2095
399CysProPheCys.....AlaAspAsp... 405
||| |||
2096 AGUACACAUAGCAAGCAUGGCGGUGGUGGUGGCGGCGGAGUAGUC 2145
406 SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
||| :::|:::|
2146 ACAGGUUGGAGUAGUCCCAUGAUCUUCUGAUGGGGUGGAGCCUGA 2195

418 ...LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeuAspPheV 433
 2196 ACCCAUUGCAGCUGUGUGUUUGUGCCUUGGCCUCCUGGCACCCUGU 2245
 433 alSerMetLeuValValThrHisPheArgLysAlaLysSerIleArgAla 449
 2246 UUGUACUGUAGUCUUCACAUUUACCGGUAUACACACAGUAGUCAAGUC 2295
 450 SerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTy 466
 2296 UCAACAGGGAACUCUGCUACAUUUUCCUGUGGCAUCUGCCGGGCUA 2345
 466 rPheProValValIleLeuTyPheGluProSerThrPheArgCysIleL 483
 2346 CUUAUGAUUUUGCCUUCUUGGAAAGGCCCAACAGAUUACUGCUACC 2395
 483 euLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyGlyThrVal 499
 2396 UUCAGAGAUGGCAUUGUGUCUCCCGCAGCAUGACUACAGGCCUU 2445
 500 ThrLeuLysLeuHisArgValLeuLysVal..... 509
 2446 GUACAAGACACACCGAUUGGAAUGGAGUCCUGGCGCAGCAAGAA 2495
 510PheLeuSerArgThrAlaGlnArgIleP 519
 2496 GAUCUGUACCAAAAGCCAGAUUCAGUGGUCUGGCCAG..... 2538
 519 roTyrMetThrGlyGlyArgValMetArgMetLeuLaValIleLeuLeu 535
 2539CUA 2541
 536 ValValPheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLe 552
 2542 GUGAUUGGUCAUUCUCAUA.....UGCAUCCAGUU 2573
 552 uGluLysGlnIleSerLeu.....IleGlyGlnG 562
 2574 GGGCAUUGGUGCCCUUUUAUUGAGGCCUCCUGACAUAAUGCAUG 2623
 562 LyLysThrSerAspHisLeuIlePheAsnMetCysLeuIleAspArgTrp 578
 2624 ACUACCCAAAGCAUUGGAGAAGUCUACCGAUUGUAAACACCAACCA 2673
 579 AspTyrMetThrAlaValAla...GluPheLeuPheLeuLeuTrpGlyVa 594
 2674 GGAGUUGUCACUCCAUUGGAAUACAGGAUUGUUAUUUGAGCUGCAG 2723
 594 lTyLeuCysTyrAlaValArgThrValProSerAlaPheHisGluProA 611
 2724 CUUCUAUGGUUACAGACCAAGAUUUGUCCAGCUACUUAACAGAGGCCA 2773
 611 rTyTyrMetAlaValAlaValHisAsnGluLeuIleSerIleAlaIlePhe 627
 2774 AGUAUUGGCCUUCACAAUGUACAGCACCGCAUUAUUGGCUAGCUUU 2823
 628 HisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTrpMetLe 644
 2824 GUGCCAUCUCACUU.....GGCAGAACUACAAA 2855
 644 uMetLeuTyrPheAlaHisThrHisLeuThrValThrValThrIleGlyL 661
 2856 CAUCACCAUGUUUUCGGUACCCUACUGGCCACAGUGGCCUAGGCU 2905
 661 euLeuLeuIleProLys.....PheSerHisSerSerAsnAsn 673
 2906 GCAUGUUUGCCGAAGGUGUACAUCAUCCUGGCCAAACACAGAGAAAC 2955
 674 ProArgAspAspIleAlaThrGlu...AlaTyGluAspGluLeuAspMe 689
 2956 GUGCGACGCGCUCACCAUACUACUGGCGGCGAUGCAUGUAGGGGA 3005

689 tGlyArgSerGlySerTyrLeu.....AsnSerSerIleAsnSerAlaT 704
 |||:||||:||||
 UGCGAAGUCAUCCUCCGACCCAGCAGACAGAUCCACGACCAUAGUACCGU 3055
 704 rpSerGluHisSerLeuAspProGluAspIleArgAspGluLeuLysLys 720
 |||:||||:||||
 GGAAGAGAGGGGCUCCUUGGGGAACACCUUAAAGGUACAAGACAGAGAGA 3105
 721 Leu.....TyrAlaGlnLeuGluIleTyrTyrArgLysLysMetI 734
 |||:||||:||||
 CUGGCCAGCACAAGUCGGAUAUAGAGUUGUUAUCCACCCCAAGGGAGUAU 3155
 734 e.....ThrAsnAsnProHisLeuG 741
 :||||:||||:||||
 GGGGAUUGGGGAGAGACAACAUGAGCAGAUCCAAUGAAAAAUCCGUGCA 3205
 741 LnlLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArg 757
 :||||:||||:||||
 CGUGGGCCAGAUAGAGAGCAGCGGGGGGAGCAGCACCUUGUGGCAGCGC 3255
 758 IleThr.....GluIlePro..... 762
 :||||:||||:||||
 CUGUCCAUAUCCAUACAAGAAGAAACCCCAACCAACGCGCCGUAU 3305
 763GluThrValSerArgGlnCysSerLysGluAspL 774
 :||||:||||:||||
 CAAGCCCUUCCCCAAGAGACCGAGAGCCGUGGCCUGGCGUGGCGUGG 3355
 774 ySGluGlyAlaAspHisGly.....ThrAlaLysGlyThr 785
 |||:||||:||||
 GCGCAGCGGGAGCGUGGGGGCGGGGGGCCACGGGCGUGGCGGCGUGC 3405
 786 AlaLeuIleArgLysAsnProGluSerSerGlyAsnThrGlyLys 802
 |||:||||:||||
 GCAGGCGCGGCCCGCCGAGUCCCGGAGAGCCGAGCGCGGCCCAAGGC 3455
 802 rLysGluGluThrLeuLysAsnArgValPheSerLeuLysLysSerHis 819
 :||||:||||:||||
 3456 G..... 3456
 819 erThrTyrAspHisValArgAspGlnThrGluGluSerSerSerLeuPro 835
 :||||:||||:||||
 3457CUGUAUGUUGGGCGGAGCGUGAGCAGCAGACUCCCG 3492
 836 ThrGluSerGlnGluGluGluThrThrGluAsnSerThrLeuGluSerLe 852
 :||||:||||:||||
 3493 GCGCCCGCGCGCGCGCUACCGUCGCGCCCAUCAGCAGCAGCAGCAGCCG 3542
 852 uSerGlyLysLysLeuThrGlnLysLysLysGluAspSerGluAlaGlu 869
 :||||:||||:||||
 3543 CGGGGG.....UCGGCCAGCGCGACGACGACGAGUGCGUG 3580
 869 erThrGluSerValProLeuValCysLysSerAlaSerAlaHisAsnLeu 885
 :||||:||||:||||
 3581 CGUGACUCGAGCGGCGUGGGCGGAGAGCAGUCCUGCAGGCGUCCUC 3630
 886 .SerSerGluLysLysThrGlyHisProArg..ThrSerMetLeuGln 901
 :||||:||||:||||
 3631 AUGGAGCAGAUCAAGCAGUGUGGUCACCGCGUACCGGCCCAACAUCAGCA 3680
 901 sSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLeuAla 917
 :||||:||||:||||
 3681 GCUAACUCCAUAGUGUGUCCACCGCGGCCCGCCAGCCCGCGGCGGCG 3730
 918 GlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGln..... 932
 :||||:||||:||||
 3731 'CCCCGUCUGCUGCUCUACCGUGAUGCCCAAGAGAGUACGAGUGGCCAGG 3780
 933LysProLeuProLysAspLysGluThrAsnArgAsnHisSerA 947
 :||||:||||:||||
 3781 ACCAUGACGACCUUUGCCGAUAUCCAGCCUUGCCCGGCCCAUCCAGGAC 3830
 947 snSerAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnPro 963

```

3831 GGGCGGGCGCGAGCCGCGCGAGGGCGCGAGGGCGGUGGGGAGCGGGCC 3880
964 AlaGluGluProArgLysProGlnLysSerGlyIleMetLysGlnGlnAr 980
3881 GGGAGAGCCCGCGG.....CCGGUCCCGAG 3906
980 gValAsnProThrThrAlaAsnSerAspLeuAsnProGlyThrThrGlnM 997
3907 GUGCGGGCG.....CCAGCCAGACC.... 3928
997 etLysAspAsnPheAspIleGlyGluValCysProTrpGluValTyrAsp 1013
3929 .....UGGAGGAGCGUGGUGG 3943
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValSe 1030
3944 CUCUCACCCCGCGGCCCCUCAGAGACUGGUGGACUGGGGAGCAC 3993
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys.... 1045
3994 ACCCCCAACUGCGAGUGCCGAGUGCGGCCUCUGUAUCCGUGUCUCC 4043
1046 .....GluLysSerHisLysProLysAla 1054
4044 CAAUAUGACACUCUUAUAGAGAUUACACUCAGAGCUCUGCGCCG 4093
1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVa 1071
4094 UGUGAA..... 4099
1071 lCysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLysLeuL 1088
4100 .UGUCCUGGAAAGCACGC..... 4117
1088 euIleSerLysThrProValLeuProGluArgAlaLysGluGluAsnGly 1104
4118 .....CGGCGUGCGGUGCG..... 4132
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
4133 .....GAGCGGAGCC 4143
1121 oProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleG 1138
4144 CCGG.....UGUUCACACACACAAU.....G 4166
1138 lHisGlnGluLysLysThrSerSerSerGluGluAsnValArgGlySer 1154
4167 GCAAGCAUAGCCGCGUGUUGCGGCCAGGGGAGGUGGCCAAGGGCACC 4216

```

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-072-574-7

seq_documentation_block:

```

; Sequence 7, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,574
; FILING DATE: 19930604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3912
; OTHER INFORMATION: /product= "HUMAN MGLUR5A"
; US-08-072-574-7

```

```

alignment_scores:
  Quality: 208.00      Length: 1084
  Ratio: 0.447        Gaps: 42
  Percent Similarity: 42.897  Percent Identity: 18.266

```

alignment_block:

```

US-09-775-181-2 x US-08-072-574-7 ..
Align seg 1/1 to: US-08-072-574-7 from: 1 to: 4085

219 GluThrGluTrpPheHisGlyLeuArgLysTrpArgProHisLeuHi 235
    ::::::::::::::::::::|||
1366 GATGTCAGTGGTTGATGATTATTCGAAGCTCCGCCAGAACAAA 1415
    ::::::::::::::::::::|||
235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgA 252
    ::::::::::::::::::::|||
1416 CCACCGAAACCCCT.....TGGTTTCAAGAAATTTTGGCAGC 1450
    ::::::::::::::::::::|||
252 tGlyAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerProPro 268
    ::::::::::::::::::::|||
1451 ATCGT.....TTTCAGTGGCGACTGGAA 1473
    ::::::::::::::::::::|||
269 TyrLeuGluCysGluAsnGlySerTyr..... 277
    ::::::::::::::::::::|||
1474 CGGTTTCCACAGGAGAACAGCAATACAAACAGACTTCAATAGTTCTCT 1523
    ::::::::::::::::::::|||
278 .....LysProGlyTrpLeuVal. 283
    ::::::::::::::::::::|||
1524 GACTCTGAAAACACATCATGTTTCAGGATTCCAAAATGGGATTGTGATCA 1573
    ::::::::::::::::::::|||
284 ...ThrLeuSerSerAlaIleTyrGlyLeu.....GlnProAsnLeu 296
    ::::::::::::::::::::|||
1574 AGCCCATCTATTTCGATGGCTGATGGCTCCACACATGCAGATGTCCTTC 1623
    ::::::::::::::::::::|||
297 ValProGluPheArgGlyVal.....Me 304
    ::::::::::::::::::::|||
1624 TGCCCGAGCTATGCAGGACTCTGTGATGCCATGAAGCCAAATTGATGGAG 1673
    ::::::::::::::::::::|||
304 tLysValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSera 321
    ::::::::::::::::::::|||
1674 GAACTTTTGGAGTCCCTGATGAAACCAATTTTACTGGGGTTTCTGGAG 1723
    ::::::::::::::::::::|||
321 spGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCys 337
    ::::::::::::::::::::|||
1724 ATACGATCCTATTTCGATGAGATGGAGACTCTCCAGGAAGGATGAATA 1773
    ::::::::::::::::::::|||

```

338	MetProIleLeuYsGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCy	354
1774	ATCAATTTCAGGAATGGGA.....	1794
354	slysAlaGlyPheTyrHisProGlyValLeuProValAsnaspPheArgA	371
1795	AAAGATTACTTTGATTATATCAACGTTGGGAAGTTGGCAACAATGGGAAT	1843
371	rgArgGlyProAspGlnHisIle.....SerGlySerThrLys	383
1844	TAAATATGATGATGATGAAGTATGGTCCAAAGAAAGCAACATCATCAGA	1893
384	AspValSerGluGlu.....	388
1894	TCGTGTGCAGTGAACCATGTGCAGAAAGCCAGATCAAGGTGATCCGAAA	1943
389AlaTyrValCysLeuProCysArgGluGly.....	398
1944	GGGAGAAAGTCAGCTGTTGTGGACCTGTACACCTTGTAAGAGAGAATGAGT	1993
399CysProPheCys.....	402
1994	ATGTCCTTGTATGAGTACATGTCAAGGCATGCCAAGCTGGGTCTGTGCCCC	2043
403	AlaAspAsp...SerProCysPheValGlnGluAspLysTyrLeuArg...	417
2044	ACTGATGATCTCACAGGTTGTCACTTGATCCGATCAGTATCTTCGATG	2093
418LeuAlaIleIleSerPheGlnGlyLeuCysMetL	429
2094	GGGTGACCTGAACCCATTCAGCTGTGGTGTGCTGCTGCTGGCCTCC	2143
429	euleuAspPheValSerMetLeuValValTyrHisPheArgLysAlaLys	445
2144	TGGCACCCTGTTGTTGTTACTGTAGCTTCATCATTTACCGTGATACACCA	2193
446	SerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySe	462
2194	GTAGTCAAGTCCTCAAGCAGGAACTCTGCTACATTATCCTTCTGGCAT	2243
462	rLeuLeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrP	479
2244	CTGCCTGGGCTACTTATGACTCTTGCTCATTTGCGAAGCCCAACAGA	2293
479	heaArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrVal	495
2294	TTTACTGTACTCTCAGAAATGGCATTTGCTCTCTCCCGCCATGAGC	2343
496	TyrGlyThrValThrLeuLysLeuHisArgValLeuLysVal.....	509
2344	TACTCAGCCCTTGTAACAACACCAACCGTATTCACAGGATCTCGGTGG	2393
510PheLeuSerArgThrA	515
2394	CAGCAAGAAGAAGATCTGACCCCAAGCCAGATTCATGAGTGCCCTGTG	2443
515	laGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAla	531
2444	CCGAG.....	2448
532	ValIleLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVa	548
2449	CTAGTGATGGCTTTCAATCTCAT.....	2472
548	lCysGlnAsnLeuGluLysGlnIleSerLeu.....	558
2473	TGCATCCAGTTGGGCATCATCGTTGGCTCTTTTAAATGGAGCCCTCGT	2521
559	..lIleGlyGlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeu	574
2522	ACATATGCATGACTACCCACAGCATTCGAGAAGCTTACCTGATCTGTAACT	2571

[illegible]

```
3349 ACCACCGCGCGGC.....TCGGCCAGCGGACGAGCAGA 3386
866 uAlaGluSerThrGluSerValProLeuValCysLysSerAlaSerAlaH 883
3387 TGTGCGTGTGCTGCACTCGGACCTGTGGCGGACGAGCTCTCTG.... 3432
883 isAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMetLeu 899
3433 .....CAGGGCTCCCTCATG 3447
900 GlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuGlyLe 916
3448 GAGCAGATCAGCAGTGTGTACCCGCTTC..... 3477
916 uAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSerGlnL 933
3478 .....ACGGCCAAACATCAGCGAG..... 3495
933 ysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSerAsp 949
3496 .....CTCAACTCCCATG 3507
950 AsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAlaGluG 966
3508 ATGCTGTCCACCGCGCGCCCGCCGCGGCTGCGCGCGCGCTCTGCTC 3557
966 luProArg...LysProGlnLysSerGly..... 974
3558 GTCCCTACCTGATCCCAAGAGATCCAGTGTCCCAAGACCATGACGAGCT 3607
975 IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspLeuAs 991
3608 TTGCGGAAATCAGCGCTCTGCGGCGCATCGAAGTCACGGCGGCGCTCAG 3657
991 pPro.....GlyThrThrClnMetLysAspAsn 1001
3658 CCGCGCGGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3707
1001 heAspIleGlyGluValCysPro.....TrpGluValTyrAsp 1013
3708 GCGCGGTCCCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3757
1014 LeuThrProGlyProValProSerGluSerLysValGlnLysHisValSe 1030
3758 CTCTACCGCGCGCTCCCGCTTCAGAGACTCGGTGGACTCGGGGAGCACA 3807
1030 rIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys.... 1045
3808 ACCCCCAACTCGCCAGTGTCCGAGTGGGCGCTCTGTATCCCGTCTCTCC 3857
1046 .....GluLysSerHisHisLysProLysAla 1054
3858 CAATATGACACTCTATCATAGAGATTACACTCAGAGCTCTCTCGTCT 3907
1055 AlaGluValCysGlnGlnSerAsnGlnLysArgIleAspLysAlaGluVa 1071
3908 TGTGAA..... 3913
1071 lCysLeuTrpGluSerGlnGlyGlnSerIleLeuGluAspGluLysLeuL 1088
3914 .TGTCCCTGGAAGACGCG..... 3931
1088 euIleSerLysThrProValLeuProGluArgAlaLysGluAsnGly 1104
3932 .....CGGCTGCGCGTGC..... 3946
1105 GlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluLeuPr 1121
3947 .....GAGCGGAGCC 3957
1121 oProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGlnIleG 1138
||||| .....
```

```
3958 CCCG.....TGTTCCACACACACAAT.....G 3980
1138 lYHisGlnGluLysLysThrSerSerGluGluAsnValArgGlySer 1154
3981 GCAGCATAGTCGCTGTTACGGCCAGGGGAGAGATGCCAAGGGCACC 4030
seq_name: /cn2_5/ptodata/2/ina/5A_COMB.seq:US-08-486-270-9
seq_documentation_block:
; Sequence 9, Application US/08486270
; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9192
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..4008
; OTHER INFORMATION: /product= "HUMAN MGLUR5B"
; OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
; insertion between nucleotides 2998 and 2999."
US-08-486-270-9
alignment_scores:
Quality: 208.00 Length: 1102
Ratio: 0.443 Caps: 43
Percent Similarity: 42.650 Percent Identity: 18.240
alignment_block:
US-09-775-181-2 x US-08-486-270-9 ..
Align seg 1/1 to: US-08-486-270-9 from: 1 to: 4181
```


3004 AAGACAGGAGACTGCCCCAGCACAGTCCGGAATAGAGTGTTTCACCCC 3053
 730 gylslysmetile.....ThrAsnA 737
 3054 CAAAGGAGTATGGGAATGGTGGGAGACAACATGACAGCTTCCAATG 3103
 737 snProHisLeuGlnLysLysArgCysSerLysLysGlyLeuGlyArgSer 753
 3104 GAAATCCGTACGTGGGCCAGAAATGAGAAGACGACCGGGGGAGCAC 3153
 754 IleMetArgArgIleThr.....Gluilepro..... 762
 3154 CTGTGGCAGCGCTGTCCATCCATCACACAGAAGAAACCCCAACCA 3203
 763 770
 3204 AACGGCGGTCAACAGCCCTTCCCAAGACACGAGAGCCGTGGGCTGG 3253
 770 erLysGluAspLysGluGlyAlaAspHisGly.....Thr 781
 3254 GCGCTGGCGCTGGCGAGCGGGAGCGCTGGGGCGGTGGGGGCCACGGGC 3303
 782 AlaLysGlyThrAlaLeuIleArgLysAsnProProGluSerSerGlyAs 798
 3304 GTGCGGGCTGGCAGCGCCCGCCAGCGGGCGCGAGTCCCCAGACGC 3353
 798 nThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeuL 815
 3354 CGGCCCAAGGCG..... 3366
 815 yLysSerHisSerThrTyrAspHisValArgAspGlnThrGluLuser 831
 3367CTGTATCATGTGGCGGAGGCTGAG 3390
 832 SerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerTh 848
 3391 GAGCACTTCCCGCGCGCGCGCGCGCGCTCCGTCGCCCATCAGCAC 3440
 848 rLeuGluSerLeuSerLysLysLysLeuThrGlnLysLeuLysGluAspS 865
 3441 GCTGAGCCACCGCGGGGC.....TCGGCCAGCCGACGGAGC 3478
 865 erGluAlaGluSerThrGluSerValProLeuValCysLysSerAlaSer 881
 3479 ACGATGTGCGCTGCTGCACTCGGAGCCTGTGGCGCGCAGCAGCTCTCG 3528
 882 AlaHisAsnLeuSerSerGluLysLysThrGlyHisProArgThrSerMe 898
 3529CAGGGCTCCCT 3539
 898 tLeuGlnLysSerLeuSerValIleAlaSerAlaLysGluLysThrLeuG 915
 3540 CATGGAGCAGATCAGCAGTGTGGTCACCGGCTTC..... 3573
 915 lLeuAlaGlyLysThrGlnThrAlaGlyValGluGluArgThrLysSer 931
 3574ACGGCCAACATCAGGGAG..... 3591
 932 GlnLysProLeuProLysAspLysGluThrAsnArgAsnHisSerAsnSe 948
 3592CTCAACTC 3599
 948 rAspAsnThrGluThrLysAspProAlaProGlnAsnSerAsnProAla 964
 3600 CATGATGCTGTCCACCGGGCCCCCAGCGCCCGCGCTCGGGCCCGCTCT 3649
 965 GluGluProArg...LysProGlnLysSerGly..... 974
 3650 GCTCTCTTACCTGATCCCAAGAGATCCAGTTGCCACACCATGACG 3699
 975IleMetLysGlnGlnArgValAsnProThrThrAlaAsnSerAspL 990
 3700 ACCTTTGCCGAATCCAGCTCTCGCGCCATCCGAAGTCACGGGGCGGC 3749

990 euAsnPro.....GlyThrThrGlnMetLysAsp 999
 3750 TCAGCCCGCGGAGGCGGCGGCGGTGGGACACGGCGCGGAGAGCC 3799
 1000 AsnPheAspIleGlyGluValCysPro.....TrpGluValTy 1012
 3800 CGCGCGCGGTCCCGAGGCTGGCGGCCCAAGCAGACCTGGAGGAGCTG 3849
 1012 rAspLeuThrProGlyProValProSerGluSerLysValGlnLysHisV 1029
 3850 GTGGGTCTCACCCCGCGCTCCCTTCAGAGACTCGTGGACTCGGGAG 3899
 1029 alSerIleValAlaSerGluMetGluLysAsnProThrPheSerLeuLys 1045
 3900 CACAACCCCAACTCGCCAGTGTCCGAGTCGGCCCTCTGTATCCGCTGT 3949
 1046GluLysSerHisHisLysProLy 1053
 3950 CTCCTCAATATGACACTCTTATCATAGAGATTACACTCAGAGCTCCTCG 3999
 1053 sAlaAlaGluValCysGlnSerAsnGlnLysArgIleAspLysAlaG 1070
 4000 TCGTTGTGAA..... 4009
 1070 luValcysLeuTrpGluSerGlnGlnSerIleLeuGluAspGluLys 1086
 4010TGTCCTTGAAAGCACGC..... 4027
 1087 LeuLeuIleSerLysThrProValLeuProGluArgAlaLysGluLys 1103
 4028CGGCTGCGCGTGCG..... 4042
 1133 nGlyGlyGlnProArgAlaAlaAsnValCysAlaGlyGlnSerGluGluL 1120
 4043GAGCGGA 4049
 1120 euProProLysAlaValAlaSerLysThrGluAsnGluAsnLeuAsnGln 1136
 4050 GCCCCCCG.....TGTTCCACACACACACAAT..... 4075
 1137 IleGlyHisGlnGluLysLysThrSerSerSerGluGluAsnValArgL 1153
 4076 ...GGCAAGCATAGTCGCTGTTACGGCCCGGAGGGAAGATGCCAAGGG 4122
 1153 ySer 1154
 4123 CACC 4126

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:01:42 ; Search time 40.53 Seconds
(without alignments)
2283.545 Million cell updates/sec

Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGAWAYPLLCILLIAQLGLG.....LSANKIAPKKEIWDSEFKV 1215

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	279.5	4.4	433	2 T21989	hypothetical prote
2	225	3.5	5327	2 T13564	microtubule-associ
3	220.5	3.5	1199	2 A41939	G protein-coupled
4	210	3.3	1171	2 A42916	metabotropic gluta
5	210	3.3	1180	2 JC2132	metabotropic gluta
6	204.5	3.2	6842	2 T29757	neurofilament prot
7	201.5	3.2	1200	2 A46194	protein UNC-89 - C
8	199.5	3.1	1212	2 JC2131	metabotropic gluta
9	194.5	3.0	3488	2 T34418	hypothetical prote
10	190.5	3.0	734	2 B42680	nucleolus-cytoplas
11	190	3.0	1390	2 T14004	trfa protein - sli
12	188	2.9	1072	2 A86827	hypothetical prote
13	187	2.9	1046	2 T29776	hypothetical prote
14	186	2.9	522	2 C96608	hypothetical prote
15	183	2.9	852	2 T06310	hypothetical prote
16	183	2.9	1560	2 T42727	proliferation pote
17	181	2.8	1358	2 A29360	SIR4 protein - yea
18	179.5	2.8	1230	2 T22458	hypothetical prote
19	178.5	2.8	1020	2 T29108	hypothetical prote
20	177.5	2.8	729	2 S68191	triadin - huma
21	177.5	2.8	3924	2 S37431	ankyrin 2, neurona
22	177	2.8	532	2 T06029	hypothetical prote
23	177	2.8	667	2 A40713	cylicin I - bovine
24	177	2.8	1165	2 T16420	hypothetical prote
25	176	2.8	2094	2 S33124	tpr protein - huma
26	175.5	2.7	1837	2 T41023	probable nuclear p
27	174.5	2.7	1131	2 A49393	activator 1 large
28	174	2.7	1020	1 OFRUH	neurofilament trip
29	174	2.7	1558	2 T29253	hypothetical prote

30	173	2.7	919	2 F81998	ribonuclease E (EC
31	173	2.7	990	2 I51618	nucleolar phosphop
32	172.5	2.7	706	2 A45990	junctional sarcopl
33	172.5	2.7	1079	2 I59362	calcium/polyvalent
34	172.5	2.7	1218	2 S71376	glutamate receptor
35	172	2.7	1310	2 T40135	probable involveme
36	171.5	2.7	971	2 T19431	hypothetical prote
37	171	2.7	1359	2 T34036	hypothetical prote
38	170.5	2.7	879	2 JC7160	metabotropic gluta
39	170	2.7	904	2 T22457	hypothetical prote
40	170	2.7	2218	2 B84683	hypothetical prote
41	169.5	2.7	598	2 B40713	cylicin I - human
42	169.5	2.7	916	2 A27864	neurofilament trip
43	169.5	2.7	919	2 F81225	ribonuclease E NM
44	169.5	2.7	2722	2 T20532	hypothetical prote
45	169.5	2.7	5138	2 B96695	hypothetical prote

ALIGNMENTS

RESULT 1

T21989

hypothetical protein F39B2.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T21989

R:Dobson, R.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19498

A:Accession: T21989

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-433 <WIL>

A:Cross-references: EMBL:Z92834; PIDN:CAB07388.1; GSPDB:GN00019; CESP:F39B2.8

A:Experimental source: clone F39B2

C:Genetics:

A:Gene: CESP:F39B2.8

A:Map position: 1

A:Introns: 137/2; 249/2; 287/3; 338/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F39B2.8

Query Match 4.4%; Score 279.5; DB 2; Length 433;
Best Local Similarity 28.7%; Pred. No. 2.7e-08;
Matches 102; Conservative 57; Mismatches 129; Indels 67; Gaps 13;

QY	456	ETILFGSLLLYFPVVILY-----PEP-----STFRCL---	LRWRLGTFATVYG	497
DB	70	EHFLSTAVALKFPRLLQLAHFFANHSKDFDQLLDKSSQIIPFDVVSFARRTHFYGLQL	129	
QY	498	TVTTLKLRVILKVFSLTAQRIPYMTGGRVYMRMLAVILLVVFVFLIGTWSSVCQNLKQIS	557	
DB	130	NCTGNHNRNLQEVYRKAQHVSVREQDMLKYLAAMLATITGLMAWT-----	176	
QY	558	LIGQKGTSDHLIFN-----MCLIDRWDMYMTAAVEFLFMGVLYCYAVRTVPVSFAHPRYM	613	
DB	177	---VGSWGDALTWRTPQCLMQGWHVWHGYELLFLLYAVRLCYKARN--SDWLERWQF	231	
QY	614	AVAVHNELIISAIFHTIRFVLASRLQSDWMLMYFAHHTLTVTVTGLLIPKFSHSSNN	673	
DB	232	TVAVCLEAVITLMANIRYSIRNSGRADTLFIVSFVHLQLTVSQVNIIVAPKFIYLSNGE	291	
QY	674	PRDDIAEAYEDELDMGRSG-----SYLNSSINSANWSE-HSLQPEDIRDELKKL	721	
DB	292	PSRRSMTLG-----GHSGRAHPHSLAKLRNLLNLTIDFAEVPIDMNPEDIRAEKRV	344	
QY	722	YAOLETYKRKKMTNNPHLQKRCSSKGLGRS---IMRRITEIPETVSQCSKDK	774	
DB	345	YTOLRWYKLNLYQDNPHISKRGKKWSKDKNKTARRIS-IP-----SCSPQTK	393	

RESULT 2

Tl3564
microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: Tl3564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL data Library, April 1999
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A:Reference number: Z17689
A:Accession: Tl3564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: *Drosophila* 576K microtubule-associated protein homolog

```

Query Match      3.5%; Score 225; DB 2; Length 5327;
Best Local Similarity 21.5%; Pred. No. 0.00087;
Matches 128; Conservative 91; Mismatches 277; Indels 100; Gaps 22;

QY 668 SHSNPNRDDTIATAYEDLDMGSGSYLNSINSINSEHSLDP-----EDIRDLKKLYA 723
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1716 AESAKGADDLKLRSRPSTTQSEAGSIKDEKSPLASEASRPSVAESYKDEAKESKE 1775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 724 QLEIYKKKMTNPHLOKKRCSKKGLGRSIMRRIETPIFVSQCSEKDEKADHGCTAK 783
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1776 E-----SRRESVAESKPLPSKEAS-----RPASVAESIKDEAKESKEESRRESVAE 1821
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 784 GTALIRK--NPPES-----SGNTOKSKKEETLKNRVF-----SLKKSHSTYDHWRDQT 828
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1822 KSLPLSKEASRPASVAESIKDEAEKSKESRRESVAESKPLPSKEASRPASVAESIKDEA 1881
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 829 EESSSLTESQEEETTENSILES-----SGKLLTQKLKEDSEAS--TESVPLVC 877
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1882 EKSKE---ESRRESVAESKPLPSKEASRPASVAESIKDEAKESKEESRRESVAESKPLPS 1938
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 878 KSAS-----AHLNISSE---RKTGHPRTSMLOKS-----LSVTASAKEKTLGLAGKTOTA 923
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1939 KEASRPASVAESIKDEAEKSKESRRESVAESKPLPSKEASRPASVAESIKDEAEKSKEE 1998
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 924 GVEERTKSQKPLPKDKETNRNHSNDNTETKDPAPQNSNPAEERPKPKQSGIMKQQRVNP 983
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1999 SRRESVAESKPLP--SKEASRPASVAES--IKDEAEKSKESRRESVAESKPLPSKEASRP 2055
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 984 TTANSLNPGTTQMKDNFDIGECVPWEYDLTGPVPSPSEKVKQHVSIIVASEMKNPTFS 1043
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2056 ASVAESIKDEAEKSK-----ESRRESAAEKSPPLPSKEASRPASVAESVKDEADKSKES 2110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1044 LKEKSHHKPKAAEV--COOSNOKRIDKAEVCLWESQSGOILEDEKLLISKTPLVPERAKEE 1102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2111 RRESMAESGKAQSIKQDQSLPKVSRP-----ESVAESVKDD-----PVKSKEP 2154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1103 NGGOPRAANYCA--GQSEELP-----PKAVASKTENENLNQIGHQEKTSISEENVNR 1152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2155 SRRESVAGSVTADSRDQPSLESKGASRPESVVDYVKDEAKQESRRESKTESVIPPKA 2214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1153 GSYNSSNNFOQLTSRAEVCPEWFETTAQPNAGRSVALPASSALSANKIAGPRKEE 1208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2215 KDKSPKEVLQPV--SMTETRETADQPMKPSQAES---RRSTAESIKASSPRDEK 2266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      3
A41939
G protein-coupled glutamate receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A41939; S15362

```

RESULT

G protein-coupled glutamate receptor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 04-Mar-1993 #sequence_revision 18-
C/Accession: A41939; S15362

R:Houamed, K.M.; Kuijper, J.L.; Gilbert, T.L.; Haldeman, B.A.; O'Hara, P.J.; Mulvihill, Science 252, 1318-1321, 1991

A:Title: Cloning, expression, and gene structure of a G protein-coupled glutamate receptor

A:Reference number: A41939; MUID:92022526

A:Accession: A41939

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1199 <HOU>

A:Cross-references: GB:M61099; NID:g397806; PIDN:AAA19497.1; PID:g204460

A:Experimental source: cerebellum

A:Note: Sequence extracted from NCBI backbone (NCBIP:60785)

R:Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi, S. Nature 349, 760-765, 1991

A:Title: Sequence and expression of a metabotropic glutamate receptor.

A:Reference number: S15362; MUID:91156047

A:Accession: S15362

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1199 <MAS>

A:Cross-references: EMBL:X57569; NID:g56646; PIDN:CAA40799.1; PID:g56647

C:Keywords: G protein-coupled receptor; transmembrane protein

		Query Match	3.5%; Score 220.5; DB 2;	Length 1199;
		Best Local Similarity	18.6%; Pred. No. 0.0002;	
		Matches 163; Conservative	120; Mismatches 330; Indels	265; Gaps
QY	357	GFYPHGVLVNNRRRGGPDHISGTSKDVSEE-----AYVCLPCREG-	398	
DB	498	GTWHGEVLNIDDDYKIO---WKSGMVRVCSEPLKGQIKVIRKEGVSCWCITACKENE	554	
QY	399	-----CPCE-----ADDSPCFQVEDKYLR-----LAIISFOGLCMLLDVFVSLVV	438	
DB	555	FVQDEFTCRACDLGWPNAEILTGCPIPVRYLEWSDIESIIATFSCIGLVTLFTLIF	614	
QY	439	YHPRKAISIIRASGILILETTILFGSLLLYFPVVIYEPSTFRICILLRWARLGATVYGT	498	
DB	615	VLYRDTVPVKSSRELCCYIILAGIFGYCGPFTLIAKPFTTTCYLQRLLVGLSAMCYSA	674	
QY	499	VTUKLHRVLKV-----FLSRTAQRIPYMTGGVMRLMAVILLVVFWELIGW	544	
DB	675	LVTNTNRIARILAGSKKKICTRPMSAWAQVLIASILSVQLTLVTVLII-----	726	
QY	545	TSSVCQNLEKOISLIGOGKTSDDLIFNMCLIDRDYMTAVA-BFLFLLMGVYLCYAVRTV	603	
DB	727	-----MEPPMPILSYPSIKE--VYLICNTSNLVGAPVGYNGLLIMSCYYAFATRNV	777	
QY	604	PSAFHEPRYNAAVAHNELIIISAIFHTIRFVLAWSLOSDMMMLYFAHTHLTVTVTIGLL	663	
DB	778	PANFNKAKYIAFTMYTCTIILWAFVIYP-----GSNYKIITTCFVLSUTVIALGCMF	831	
QY	664	IPK-----FSHSSNNRDDIATEAYEDELDMG-----RSGYLNSINSAMSEHSLOP	711	
DB	832	TPKMYIIIAKPERNVRSAFIT---SDVVRMHVGDKLPCRSTNFLN-----	874	
QY	712	EDIRDELKULYAOLEIYKRKKMTINPHLOKKRCS-----KKGILGRSIMRRIETPE	764	
DB	875	-----IFRRKKPGAGANSNGKSVSWSEPGRGQAOPKHQWQRLSHVYKT	919	
QY	765	VSROCKEDKEGADHGCTAKTALIRKNPPSSCNTCKSEETLKNRVESLKKSHSYDHY	824	
DB	920	NETACNQ-----TAVIK--PLTKSYQSCKSLTFSD-----	948	
QY	825	RDOTESSSLPTSEBEETTENTLESGLKLTOKLKEDSEAESTESVPLVCKSASAHH	884	
DB	949	-----ASTKTYLNVEBEDNPSPAHFSPPPSPSMVVHR	980	
QY	885	LSSEKKTGHPTSMLOKSLSVIASAKEKITGLACKQTQACVEERTKSQPLPKDKETRNR	944	
DB	981	RGPVATPTPLPHL-----TAEETPLFLADSVIPKGL-----PPLPOQQ-----	102	
QY	945	HSNSDNTETKDPAQNSNPAEPRKPKQSGIMKQQRVNPTTANSDL-----NPQTTO	996	

Db 775 FTMVTCIIWLAFLPIYF-----GSNYKIITCMFVSLSATVALGCMFVPKVIYIIILAKP 828
 QY 671 SNNPRDDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLOPEDIRDELKK----- 720
 Db 829 ERNVSFAFTTSTVVRMHVGDKGSSAASRSSSLVNLWKRGGSGETLSSNGKSVTWAQNE 888
 QY 731 -----LYAQLEIYKRKKMTIINNPHLOK---KRCSSKKGLGRSIMRRITEIPETVSRQCS 770
 Db 889 KSTRGQHLWQLSVHINKENPNCTAVIKPPKSTENRG-----PCGAAAGGS 936
 QY 771 KEDREGADHGCTAKTALIRKNPPSSGNTGKSEETLKNRVFSLKKSHSTYDHRVDQTEE 830
 Db 937 GPGVAGA--GNAGCTATGGPEPPDAPKA-----LYDVAEA 970
 QY 831 SSSLPTEQSEBETTENSTLESLSGKKLTQKLKEDSEAESESVPLVCKSAHNLSEKK 890
 Db 971 EESFPAAARPSPSITLSHLAG-----SAGRTDDDAPLSHSETAARSSSQGS 102
 QY 891 TGHPRFTSMLOKSLVIAASKEKTLGLAG-----KTQTAGVEER 928
 Db 1021 LMEQISSVVTFTANISELASMMLSTATPGPPTPICSSYLIPKEIQLPTMTTFAE-- 107
 QY 929 TKSQKPLPKDRETNRNHSNDTETKDPAPQNSNPAAE-PRKPKQSGIMKQQRVNPTTAN 987
 Db 1079 ---IOPUPAIEVT---CGAAGATGVSPAQETPTGAESAPGKPDLELVALTPPSP--- 112
 QY 988 SDLNPGTQMKNDFDIGEVCPEWYDITPGVPVSESK 1024
 Db 1128 -----FRDVSQSTTPNSPVSEALCIFSSPK 1155

 RESULT : 5
 JC2132
 metabotropic glutamate receptor 5 A - human
 C:Species: Homo sapiens (man)
 C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 05-Jan-1999
 C:Accession: J02132
 R:Minakami, R.; Katsuki, F.; Yamamoto, T.; Nakamura, K.; Sugiyama, H.
 Biochem. Biophys. Res. Commun. 199, 1136-1143, 1994
 A:Title: Molecular cloning and the functional expression of two isoforms o
 A:Reference number: JC2131; MUID:94197696
 A:Accession: JC2132
 A:Molecule type: mRNA
 A:Residues: 1-1180 <MIN>
 C:Comment: This protein is coupled to quanine nucleotide binding proteins.
 C:Keywords: glycoprotein; neurotransmitter; receptor; transmembrane protei
 F:580-604/Domain: transmembrane #status predicted <TM1>
 F:617-637/Domain: transmembrane #status predicted <TM2>
 F:644-664/Domain: transmembrane #status predicted <TM3>
 F:694-714/Domain: transmembrane #status predicted <TM4>
 F:738-759/Domain: transmembrane #status predicted <TM5>
 F:773-794/Domain: transmembrane #status predicted <TM6>
 F:803-827/Domain: transmembrane #status predicted <TM7>

 Query Match 3.3%; Score 210; DB 2; Length 1180;
 Best Local Similarity 19.1%; Pred. No. 0.00076;
 Matches 179; Conservative 120; Mismatches 400; Indels 240; Gaps

 QY 219 ETEWFHGLRRKWRPHLHRRGNQCPRLGHSWR-----RKDGLGDKSHFKWSPPYLECE 273
 Db 333 DVKWFDDYYLKLRPETNHRNP-----WFQEFWQHRFOCRLEGFPQENSKYKNKT-----CN 382
 QY 274 NG-----SYRFGWLIV-TLSSAIYGL--QPNLVPEFGV-----MKVDINLQ 311
 Db 383 SSSLTKLTHHVQDSKMGFVINAIIYSMAYLNMQMSLCPGVAGLCDAMKPIDGRKLLLESLM 442
 QY 312 KVIDIQCSSDQWFGSGTHKCHLNNSECPKIGLGPVLGAYECICKAGFYHPGLVPVNNRR 371
 Db 443 KTNFTGVSGDITLFDENGSDSPGEIIMFNKEMG-----KDYFDYINVGSDWNGEL 492
 QY 372 RGPQDHI-----SGSTKDYSEE-----AYVCLPREG-----CPFC 402

Db 493 KMDDEVSKSNIIRSVCSBPCKGQIKVIRKGEVSCCWTCTCKENYVPEVDTCKAC 552
Qy 403 -----ADD-SPCFQVEDKYLR-----LAIISFOGLCMLLDVFSMLVYVYHFRKAKSIRA 449
Db 553 QLGSMPTDGLTCDLIPVQYLRWGPDPPIAAVFAACLGALLATFVTVVFIIRYDTPPVKS 612
Qy 450 SGLILLETILFCSLLLYFPVILFEPSTFCILLRWALRLLGFATVYGTIVTLKLRVLKV 509
Db 613 SSRELCTIILAGICLGYLCTCLIAKPKQIYCYLQIRIGLSPAMSYSALVTKTNRIARI 672
Qy 510 -----FLSRTAQRIPVMTGGRVWRMLAVILLVVEFLIGTSSVQCNLEKQ 555
Db 673 LAGSKKKCTCKPRMSACQ-----LVIAFIL-----CQUGII 708
Qy 556 ISL-----IQGKTSDHILFNCLIDRWDMYTAFA-EFLFLMGVLYCYAVRTPVSFAF 607
Db 709 VALFIMEPPDIMHDYPSIREVYLICNTNLGVVTPPLGYNGLLILSCTFYAFKTRNVPANF 768
Qy 608 HEPRYMAVAVHNELIISAFHTIRFVLASRLQSDWMLMYFAHHTLVTVVITGLLLIPK- 666
Db 769 NEARYIAFTMTTCIWLAFVIYF-----GSNYKIITMCFVSLSATVALGCMFVPKV 822
667 ---FSHSSNNPRDDIATE-AVEDELDGRSGSYL--NSSINSANSEHSLDPEDIRDELKK 720
Db 823 YIILAKPERNVRSAFTTSTVVRMHVGDGSSAASRSLLVNLWKRSGSGSETLRSGKS 882
Qy 721 -----LYAQLEIYKRRKMTNNPHLOK---KRCSSKGLGRSIRMRITPEI 763
Db 883 VTAQNEKSSRGQLWQLRSLHIINKENPNQTAIVKPPKSTESRGLGP-----932
Qy 764 TVSRQCKEDGADHGTAKGTALIRKNPPSSNGTNGKSKETLKNRVFSLKSHSYDH 823
Db 933 -----RGAG-GSAGVGATGAGCAGAGGPGGSPDAGPKALY-----969
Qy 824 VRDQTESSSLPTESQEBETTENTSLGSKKLTQKLKEDSEAESESVPLVCKKSASAH 883
Db 970 ---DVAEAEHPAPARPRSPSPITLSHRAG---SASRTDDVPSLHSEPVARSSSSQ 1023
Qy 884 NLSSEKKTGHPR-TSMLOKLSUSVIASAKETLGLAG-----KTQTAGVEE 927
Db 1024 SLMEQISSVTRFTANISLSELMMLSTAAPSPGVGAPLCSYLLIPKEIQLPTTTTFAE- 1082
Qy 928 RTKSQKPLPKDKETNRHNSDNTETKDPAPON--SNPAEPRPKQSGIMKQORVNPPT 985
Db 1083 ----IQPLPAIEVGGQAQAAGAAQDAARESPAAGPEAAAKPDLELVALTPPSP-- 1136
Qy 986 ANSDLNPGTTQMKDNFDIGEVCPVEVDLTGPGVPVPSSEK 1024
Db 1137 -----FRSDVDSGGTTPNSPVSESLCIPSSPK 1164

ULT 6
9757
protein UNC-89 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
A:Accession: T29757
R:Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid C09D1.
A:Reference number: 220679
A:Accession: T29757
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6642
A:Cross-references: EMBL:AF003131; PIDN:AAB541132.1; GSPDB:GN00019; CESP:unc-89
A:Experimental source: strain Bristol N2; clone C09D1
C:Genetics:
A:Gene: CESP:unc-89
A:Map position: 1
A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 3.2%; Score 204.5; DB 2; Length 6642;
Best:Local Similarity 22.4%; Pred. No. 0.016;
Matches 151; Conservative 98; Mismatches 270; Indels 155; Gaps 31;
Qy 609 EPRYMAVAVHNELIISAFHTIRFVLASRLQSDWMLMYFAHHTLVTVVITGLLLIPKFS 668
Db 1238 EPKESVYVVEKQDLSS-----SEVQKIAQQVKEASPEATTITMETSLSITSTKT 1286
Qy 669 HSSNNPR-----DDIATEAYEDELDM-----GRSGSYLNSSSINSANSE--HSLDPE-DI 714
Db 1287 TTMSTTEVTSTVGTVVETKESSESATTVTGGSGGVTEGSSISVSKIEVVSKVTDSDQDV 1346.
Qy 715 RDELKK---LYAQLEIYK-----RKKMTNNPHLOKRCCKGLGRSIRMRITPEITV 765
Db 1347 REGTPKRRVSAFEEELPEVIDSDRKKKKSPDPKKEKSPK-----TEEKPASP 1396
Qy 766 SRQCKEDGADHGTAKGTALIRKNPPSSNGTNGKSKETLKNRVFSLKSHSYDHR 825
Db 1397 TKKTGEEVKSPKSPASPT-----KKEKSPAABEVKSPTK-----EKSPSS---P 1440
Qy 826 DQTESSSLPTESQEBETTENTSLGSKKLTQKLKEDSEAESESVPLVCKKSASAHNL 885
Db 1441 TKKESPSPTKTKGDEVKEKS-----PPKSPTRKKEKSPKEDVKSPVKKKESPDATNI 1495
Qy 886 ---SSEKKTGHPRTSML-----QKSLSVIASAKETLGLAGTQTAGVEERTKSQKPLPK 937
Db 1496 VEVSSETTIETKTETMTTETMTHSEESRTSVKKEKT-----PEKVDKPKS--PTKK 1545
Qy 938 DKETNRHNSDNTETKDPAPONNP---AEEPRKPKQSGIMKQORVNPPT-----T 985
Db 1546 DKSPK-----SITEIKSPVKKESPEKVEEKSPASPTKKEKSPKSPASPTKKSENEVKSP 1601
Qy 986 ANSDLNPGTTQMKDNFDIGEVCPVEVDLTGPGVPVPSSEKVKHVSIVASEMEKNPPTSLEK 1045
Db 1602 TKKESKSEKSVVEELKSPKESPEKADDDKPSPTKKEKSPKES-----ATEVDKSPTKKEK 1657
Qy 1046 --EKSHHKPAEYCOQSNOKRID-----KAEVCLWESQGG 1079
Db 1658 SPEKVEEKPTSPTKKESPTKKTDDVKSPTKKESQTVTEKSPASPTKKE---KSPK 1713
Qy 1080 SILDEKLLISKTIVLPERAKEENGQOPRAANVCAGOSEELPPKAVA-----SKTENENLN 1135
Db 1714 SVVEEVKSPKES---PEKAE---KPKSPT-----KKEKSPESAABEVKSPTKKEKSP 1762
Qy 1136 QIGHOEKKTSSSEENVRGYSNNFQOPLTSRAEVCWPWF-ETPAOPNAGRSVALPASS 1194
Db 1763 EKSAEERKPSPTKKESSPVKMADEVKSP--TKKEKSPKVEEKASPT--KKEKTPEKS 1818
Qy 1195 ALSANKIAGPRKEE 1208
Db 1819 A--AEELKSPTKKE 1830

RESULT 7
A46194
neurofilament protein NF-220, high-molecular-weight splice form - longfin squid
C:Species: Loligo pealeii (longfin squid)
C:Date: 22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Mar-2000
C:Accession: A46194
R:Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Gainer, H.; Battey, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992
A:Title: A high-molecular-weight squid neurofilament protein contains a lamin-like ro
A:Reference number: A46194; MUID:92357751
A:Accession: A46194
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1200 <WAY>
A:Cross-references: GB:M94389; NID:g161291; PID:g161292
A:Experimental source: stellate ganglion
A:Note: sequence extracted from NCBI backbone (NCBIP:113499)
C:Superfamily: neurofilament triplet H protein
C:Keywords: alternative splicing

Query Match	3.1%	Score 199.5;	DB 2;	Length 1212;
Best Local Similarity	18.2%	Pred. No. 0.003;		


```

QY 1047 KSHHKPAAEVCQSSNQKRIDKAEVCLMESQGSILEDEKLLISK-TPVILPERAKEENG 1105
      |||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
Db 450 GSQOKPQSRKADSSSSSEESSSE---EATKKSVTTPKARVTKAAAPSLPAQAPRAGG 506
      |||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||

QY 1106 QPRAANVCAGQEE--LPKAVASR-----TENENLNOIGHOEKKTSSSEENVRSYN 1156
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 DSSSDSESSSEEEKTPPKPAKKKAAGAVPPTPKKAAAESSSSSSSE-----D 560
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1157 SSNNFOOPLTSAEVCWPEFTPPAPNAGRVSVALPASSAISANKIAGRKEE 1208
      ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
Db 561 SSEEKKPKSKA-----TP-KPQAGKANGVPASQ---NGKAGKESEE 599
      ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||

RESULT 11
Tl4004
trfA protein - slime mold (Dictyostelium discoideum)
Species: Dictyostelium discoideum
Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: Tl4004
R:Saito, J.; Adachi, H.; Sutoh, K.
J. Biol. Chem. 273, 24654-24659, 1998
A:Title: Dictyostelium TRFA homologous to yeast Ssn6 is required for normal
A:Reference number: Z17852; MUID:98406112
A:Accession: Tl4004
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-11390 <SAI>
A:Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BAA3314
C:Genetics:
A:Gene: trfA
A:Introns: 333/3; 364/3; 637/1

```

[illegible]

RESULT¹²
 A:Accession: A86827
 A:Title: hypothetical protein yqfG [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL140)
 C:Species: *Lactococcus lactis* subsp. *lactis*
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
 R:Accession: A86827
 R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
 Genome Res. In press, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium.
 A:Reference number: A86625
 A:Accession: A86827
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1072 <STO>
 A:Cross-references: GB:AE005176; NID:gi12724625; PIDN:AAK05715.1; GSPOB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: vafg

		Query Match	2.9%; Score 188; DB 2; Length 1072;
		Best Local Similarity	18.9%; Pred. No. 0.011;
		Matches	146; Conservative 109; Mismatches 264; Indels 254; Gaps 32;
Qy	618 HNELIISAIPHTIRPVLASRLQSDDMLMIFAHTHLIV-----PVITGLLL 663 : : : :	Df	176 NNNLTSLGSGNTEIMAKAFOSNOITIIFADESLTVIDSAAFSGSVSOSTLGIGVTL 235 : : : :
Qy	664 I-----PKFSSHNNPRDIAATEAYEDELDMGRGSYLNSSIN----- 702 - - - - - : : : :	Df	236 ADDVFNTSPFLFGQLSDLPDEIRT-----ISVNSSLFDKSWINDSSQSSTVSYSTENA 289 - - - - - : : : :
Qy	703 ----AWSEHSLDPEDIRDCLKL---YAOLEIYKRKKMITNPHLQKRRCKKGILGRSIM 755 : : : :	Df	290 DTPVPSSSNVEQQDVASSESTODANSASYPISEASSVTDN-----TL 333 : : : :
Qy	756 RRTEIPTETYSRQCSEKEKA-----DHCTAGKTALIKRNPPESPSCNTCK-SKETLK 808 : : : :	Df	334 NSISSLOSSISTOTSNSQGACASTAEISYDSSENSLSLSSNQINSNSEKDSQNCSLGG 393 : : : :
Qy	809 NRVSFLAKS-IHYTDHVDRDTQEESSL-----PTE-----SQEBETTENTLES 851 : : : :	Df	394 SSMSSNEESESHNSSNI-NETNNSSEITNLPPSPNETPSNVSDOTSEASTSNSSI-S 451 : : : :
Qy	852 LSGKKLVTKUKEDBEARSTE-----SYPLVCVKCSAHLN----- 885 : : : :	Df	452 LSPNISSTSDESATNSDFSNVAEVANNSSLAVNNSVVSSLSSTADNLGINQSGSD 511 : : : :
Qy	886 -----SSEKKTYCH-----PRTSMLOKSLS----- 904 : : : :	Df	512 NLTKDSGEITSYGAPLSSNOTSSEASTNSNSTSLSPNSISSTSVLESTTSSSNFSNAVE 571 : : : :
Qy	905 ---VIASAKEYKLGLAGKTOTA-----GVERTKSKOKPL-----PKDKETN 942 : : : :	Df	572 VANNSLASNSSSVLSSTTSADNLFINOGSDNLFKDSEISTCGAPLSSNOTSSEAS 631 : : : :
Qy	943 RNHNNDON-----TETKDPAQPONSANPEPRKPQSKGMKOQRYN-----PTTN 987 : : : :	Df	612 SNMWSSINSPSLSLTSNESSATNQNSSEATKVDDNNSSTHSSNILNGSNDSSDSD 691 : : : :
Qy	987 SDLPNGTTQMKNDFDIGEVC---PWEEVDLTPGP--VPSES KVKQHVSIVA SEKKNPTTF 1042 : : : :	Df	692 SDSDSLMSLLSSPNLETQTITSSKPSVENNYISENPKKYSSSNSVOENST-DHEMSTNPKS 749 : : : :
Qy	1043 SLKEKSHHKPKAAFEVCCQGNOKRIDKAEVCLWESQGOSILEDEKLLISKTPVLPERAKEE 1102 : : : :	Df	750 SIS-----SPISTTSSSOOQ-----EQSQLNLTETEII---NNPITFNNSGSE 789 : : : :
Qy	1103 NGGOPRAANYCA-----GOSEEPLPPKAVASKTENENLQIGHOEKKTSTSEENVRG SYN 1156 : : : :	Df	790 NSS----AASIITYSYNNNSSESSETGCIFYINFAORDNGSRISHSLPSSSNENN NV----- 840 : : : :

10

QY 1157 SSNNFOPLTSRAEVCWPEETPAQPNAGRSVALPASSALS-ANKIAGRKEE 1208
Db 841 SSISOAILLES-----SKSSTNKRSSLSLSINSTSHUPONED 876

RESULT 13

T29776

hypothetical protein C50F2.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29776

R.Du, Z.; Le, T.T.

submitted to the EMBL Data Library, November 1996

A:Description: The sequence of C. elegans cosmid C50F2.

A:Reference number: 220684

A:Accession: T29776

A:Status: preliminary; translated from CB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1046 <DUZ>

A:Cross-References: EMBL:U80445; PIDN:AAB37795.1; GSPDB:GN00019; CESP:C50F2.2

A:Experimental source: strain Bristol N2; clone C50F2

Genetics:

Gene: CESP:C50F2.2

Map position: 1

A:introns: 72/2; 888/1; 940/1; 986/3

Query Match 2.9%; Score 187; DB 2; Length 1046;
Best Local Similarity 22.3%; Pred. No. 0.012;
Matches 122; Conservative 86; Mismatches 220; Indels 118; Gaps 26;

QY 676 DIATEAYEDELDMGRSG-----SYLNSSINSAMSEHSLSLDPEDI-----RDELKLYAQ 724
Db 138 EDLEWLYLNQLETTNDGMPYVNVYKATNCPEKSLKESGSLLPVLEKKEKREKQEDQ 197
QY 725 -LEIYRKMMITNNPHLOKRCCKGLGHSIMRRITETPETVROCSKEDKAGADHGTAK 783
Db 198 KAVVLEKRPVTKVQDHLKRRSKSPIGRGQSDKDIISNKKRLNKREEIDE-----253
QY 784 GTALIRKNPPESSNGTGSKSEETLKNRV-FSLKKSHSTVDHVRDOTEESSSLP-TESQEE 841
Db 254 -RSK TAKAP-----ISQVQFPMPAPQAPDAVASKVRQKPVDTIPHPRP 299
QY 842 ETENSTLESGLKLTOKLKEDSEASTESVPLYCKSASAHNLSSEK-----TGHP--- 894
Db 300 SASRNSQKPVASKIQEK-KTEVRTKSSPTNPVINSQ-----PSQKKPADLTPHPKRP 353
QY 895 -----RTSMLOKSLSVIASAKEKTLGLAGKTOTAGVEERTK-SOKPLPKOKETNRN 944
Db 354 PITPETLPSRNSIQKPA--SKIQEKKTEVIARSQKS--EASTRIAQKPSPTNPVNFK 409
QY 945 HNSDNT-----TKDPAPONSNAEEP-RKPOKSGIMKQORVNPNTTANSDLNPGT 994
Db 410 QPSQPADLTPKPIPLASKSPITRSEVSTERTKRAPVIMDDPKDSTA-KELPKEL 468
QY 995 TOMKDNFDIGEYCPWEVYDLTPGVPSESKVOKHSIVASEMEKNPTPSLKEKSHHKPKA 1054
Db 469 PKAKSHFEF-KVKPIEKKDASPDKTTST-----SASPPNPDPMKQKVFVPTS 517
QY 1055 ARVCOOSNQKRDKAECVCLWESQOSILEDEKLLISKTPVLPERAKENGGOPRAANVCA 1114
Db 518 -----KKOK-----LTETKN-----KQKSTINESPV-----EQNSGONEASNTK 554
QY 1115 GQSEEL-----PPKAVASKTENENLQIGHQEKKTSSSEENVRGSYNSNNFQOPLTSRA 1169
Db 555 LKQDELNHVVPKNPEAETPD-----RQESVOKSNEASATSTKPSDPVETP--SD 605
QY 1170 EYCPWE 1175
Db 606 EYDPOE 611

RESULT 14

C96608

hypothetical protein F25P12.91 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C96608

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
Nure, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C96608

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-522 <STO>

A:Cross-References: GB:AE005173; NID:g9954748; PIDN:AAG09099.1; GSPDB:GN00141

Genetics:

Gene: F25P12.91

Map position: 1

Query Match 2.9%; Score 186; DB 2; Length 522;
Best Local Similarity 19.7%; Pred. No. 0.0054;
Matches 96; Conservative 90; Mismatches 218; Indels 84; Gaps 15;

QY 709 LDP-----EDIRDELKLYAQLEIYRKMMITNNPHLOKRCCKGLGHSIMRRITETPT 764
Db 22 LDPKEGNEVEVEVEMEVAKSIEKVAKKDESSGSKKDKKEKK--GKNVDEVAKDD 79
QY 765 VSRQ-----CSKEDKGGADHG---TAKGTALIRKNPPESSNGTGKSEETLKRVFSLKKS 817
Db 80 DKKDKGKVMYKHEBG--HGDLEVKESDVKEHEKHEKHEKHEELEEKEGKKKN 137
QY 818 HSTYDHRDOTEESLSLPTESQEEETTENTSTLESLSGKLTOKLKEDSEASTESVPLVC 877
Db 138 KKEKDESGPEEKNKADKEKHEDVSQKEELEEDGKNKKKKEDESGTEEEKKKPKKE 197
QY 878 KSASAHNLSSEKKTGHPRTSMLOKSLSVIASAKEKTLGLAGKTOTAGVEERTKSQKPLPK 937
Db 198 KKQKEESKNEK-----KVKGKEKEGKEDLEKEDEKK--KE 234
QY 938 DKETNRNHSNDNTETKDPAPONSNAEPRKPKQKSGIMKQORVNPNTTANSDLNPGTQM 997
Db 235 HDETQEMKEKDSKKNKKKDESCAEEKKKPKDEKDEKDEKDKLKGKKGKGEK 294
QY 998 KDNFDIGEYCPWEVYDLTPGVPSES-----KVKHVSIVASEMEKNPTPSL 1044
Db 295 PEKEDEGK--KTKHDATEQEMDDEADHKEGKKKKKDKAKKKTVIDEVECK-----ET 348
QY 1045 KEKSHHKPKAAEVCQOSNQKRDKAECVCLWESQOSILEDEKLLIS--KTPVL----- 1095
Db 349 KDKODDE--GETKQKNNKKEKSE-----KGEKVDKDKKENPLETEVMSRDKLE 399
QY 1096 ---PPRAKEENGOPRAANVAGQSEELPPKAVASKTENENLN-----QIGHQPKKTS 1146
Db 400 BPEAKKEEDTTEKKSKVKGSEEGKGGKKKKDKKKNKKDKTKPRKMTDEEBEKKKDS 459
QY 1147 SEENVGRS 1154
Db 460 KDVKIEGS 467

RESULT 15

T06310

hypothetical protein F11C18.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:14:27 ; Search time 24.77 Seconds
(without alignments)
1680.275 Million cell updates/sec

Title: US-09-775-181-2

Perfect score: 6382

Sequence: 1 MGAWYPELLCLLLAQLGLG.....LSANKIAPRKEIWDSEFKV 1215

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	220.5	3.5	1199	1 MGR1_RAT	P23385 rattus norv
2	211	3.3	1203	1 MGR5_RAT	P31424 rattus norv
3	199	3.1	704	1 NP41_RAT	P41777 rattus norv
4	197.5	3.1	1212	1 MGR5_HUMAN	P41594 homo sapien
5	193	3.0	1194	1 MGR1_HUMAN	Q13255 homo sapien
6	184	2.9	976	1 MGR_DROME	P91685 drosophila
7	181	2.8	1358	1 SIR4_YEAST	P11978 saccharomyc
8	178	2.8	1637	1 MRSP_STAAU	P80344 staphylococ
9	177.5	2.8	728	1 TRDN_HUMAN	Q13061 homo sapien
10	177.5	2.8	3924	1 ANK2_HUMAN	Q01484 homo sapien
11	177	2.8	667	1 CYL1_BOVIN	P35662 bos taurus
12	176	2.8	2349	1 TPR_HUMAN	P12270 homo sapien
13	175.5	2.7	1079	1 CASR_MOUSE	Q9qy96 mus musculu
14	174.5	2.7	1131	1 AC15_MOUSE	P35601 mus musculu
15	174	2.7	1020	1 NFH_HUMAN	P12036 homo sapien
16	172.5	2.7	633	1 MLH1_TETRA	P40631 tetrahymena
17	172.5	2.7	705	1 TRDN_RABIT	Q28820 oryctolagus
18	172.5	2.7	1079	1 CASR_RAT	P48442 rattus norv
19	169.5	2.7	598	1 CYL1_HUMAN	P35663 homo sapien
20	169.5	2.7	915	1 NEM_HUMAN	P07197 homo sapien
21	169	2.6	999	1 MGR1_CAEEL	Q09630 caenorhabdi
22	168.5	2.6	1233	1 YF16_YEAST	P23597 saccharomyc
23	168.5	2.6	1257	1 RBH1_HUMAN	P29374 homo sapien
24	168.5	2.6	1849	1 IGA4_HAEIN	P45386 haemophilus
25	168	2.6	1411	1 TCOF_HUMAN	Q13428 homo sapien
26	167.5	2.6	816	1 YG3A_YEAST	P53278 saccharomyc
27	167	2.6	836	1 NOT3_YEAST	P06102 saccharomyc
28	166	2.6	1523	1 SON_HUMAN	P18583 homo sapien
29	166	2.6	2375	1 ATRX_HUMAN	P46100 homo sapien
30	165.5	2.6	952	1 IF41_YEAST	P39935 saccharomyc
31	165	2.6	1275	1 TRP_DROME	P19334 drosophila
32	164.5	2.6	589	1 HIBN_XENLA	P06180 xenopus lae
33	164	2.6	2476	1 ATRX_MOUSE	Q61687 mus musculu

ALIGNMENTS

RESULT 1

ID	MGR1_RAT	STANDARD;	PRT;	1199 AA.
AC	P23385;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.			
GN	GRM1 OR MGLUR1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;			
RX	MEDLINE=91156047; PubMed=1847995;			
RA	Masu M., Tanabe Y., Tsuchida K., Shigemoto R., Nakanishi S.;			
RT	"Sequence and expression of a metabotropic glutamate receptor.";			
RL	Nature 349:760-765(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=Brain;			
RX	MEDLINE=92022526; PubMed=1656524;			
RA	Houamed K.M., Kuipier J.L., Gilbert T.L., Haldeman B.A., O'Hara P.J.,			
RA	Mulvihill E.R., Almers W., Hagen F.S.;			
RT	"Cloning, expression, and gene structure of a G protein-coupled			
RT	glutamate receptor from rat brain.";			
RL	Science 252:1318-1321(1991).			
RN	[3]			
RP	ALTERNATIVE SPLICING (ISOFORM 1B).			
RC	TISSUE=Brain;			
RX	MEDLINE=92110002; PubMed=1309649;			
RA	Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;			
RT	"A family of metabotropic glutamate receptors.";			
RL	Neuron 8:169-179(1992).			
RN	[4]			
RP	ALTERNATIVE SPLICING (ISOFORM 1C).			
RC	TISSUE=Brain;			
RX	MEDLINE=93066232; PubMed=1438218;			
RA	Pin J.-P., Waeber C., Prezeau L., Bockaert J., Heinemann S.F.;			
RT	"Alternative splicing generates metabotropic glutamate receptors			
RT	inducing different patterns of calcium release in Xenopus oocytes.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:10331-10335(1992).			
RN	[5]			
RP	FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS			
CC	MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-			
CC	CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL			
CC	ACTION OF GLUTAMATE IN THE CNS. SUCH AS LONG-TERM POTENTIATION IN			
CC	THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1A (SHOWN HERE), 1B AND 1C;			
CC	ARE PRODUCED BY ALTERNATIVE SPLICING. 1B AND 1C ARE C-TERMINALLY			
CC	TRUNCATED FORMS OF 1A.			
CC	-!- TISSUE SPECIFICITY: IS PREDOMINANTLY EXPRESSED IN CEREBELLAR			
CC	PURKINJE CELLS, CA2-CA3 PYRAMIDAL CELLS OF THE HIPPOCAMPUS, AND			
CC	MITRAL AND TUFTED CELLS OF THE OLFACTORY BULB.			

Q14832 homo sapien
P46012 caenorhabdi
P31422 rattus norv
Q02952 homo sapien
Q43493 homo sapien
P98193 rattus norv
P30414 homo sapien
P51825 homo sapien
P25386 saccharomyc
P44487 caenorhabdi
Q12263 saccharomyc
Q03661 saccharomyc

CC -!- MISCELLANEOUS: ACTIVATED BY QUISOVALATE > GLUTAMATE > IBOTENATE >
CC TRANS-1- AMINOCYCLOPENTYL-1,3-DICARBOXYLATE; INHIBITED BY
CC 2-AMINO- 3-PHOSPHONOPROPIONATE.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57569; CAA40799.1; -
CC EMBL; M61099; AAA19497.1; -
CC EMBL; S48085; AAB24138.1; -
CC PIR; S15362; S15362.
CC PIR; A41939; A41939.
CC GCRDB; GCR_0216; -
CC InterPro; IPR000162; -
CC InterPro; IPR000337; -
CC InterPro; IPR001256; -
CC InterPro; IPR001828; -
CC Pfam; PF00003; 7tm_3; 1.
CC PRINTS; PF01094; ANF_receptor; 1.
CC PRINTS; PR00248; GPCRMR.
CC PRINTS; PR00593; MTABOTROPICR.
CC PRINTS; PR01051; MTABOTROPICR.
CC PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE; PS02059; G_PROTEIN_RECEP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family; Alternative splicing.
CC SIGNAL 1 18
CC CHAIN 19 1199 METABOTROPIC GLUTAMATE RECEPTOR 1.
CC DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 593 615 I (POTENTIAL).
CC DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 630 650 II (POTENTIAL).
CC DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 662 680 III (POTENTIAL).
CC DOMAIN 681 706 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 707 727 IV (POTENTIAL).
CC DOMAIN 728 750 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 751 772 V (POTENTIAL).
CC DOMAIN 773 785 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 786 808 VI (POTENTIAL).
CC DOMAIN 809 814 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 815 840 VII (POTENTIAL).
CC DOMAIN 841 1199 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 1014 1034 GLN/PRO-RICH.
CC DOMAIN 1074 1080 GLN/PRO-RICH.
CC DOMAIN 1126 1135 ASP/GLU-RICH (ACIDIC).
CC DOMAIN 1140 1199 SER-RICH.
CC CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
CC VARSPLIC 887 906 NSNGKSVSWSEPGGRQAPK -> KKRQEPSPSQCPSAH
CC AQL (IN ISOFORM 1B).
CC VARSPLIC 907 1199 MISSING (IN ISOFORM 1B).
CC VARSPLIC 888 897 SNKGSVSWSE -> FALDRONTVY (IN ISOFORM 1C).
CC VARSPLIC 898 1199 MISSING (IN ISOFORM 1C).
CC SEQUENCE 1199 AA; 133235 MW; EEE5A04C50694B9F CRC64;

Query Match 3.5%; Score 220.5; DB 1; Length 1199;
Best Local Similarity 18.6%; Pred. No. 0.0004;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;
Qy 357 GFYHPGVLPVNNFRRRGPDQIHSGSTKDVSEE-----AYVCLPREG- 398

Db 498 GTWHEGVNLIDDDYKIQ-----MNRSGMVRVSCSEPCCLKGQIKVRKGEVSCCWTACKENE 554
Qy 399 -----CPFC-----ADSPCFVQEDKYL-----LAISFGQCLMLDLDFVSLWV 438
Db 555 FVQDEFTCRADLGMWPAELTGCEPIPVIRLEWSDIESIIAIAFSCILGILVTLFVTLIF 614
Qy 439 YHFRKAKIRASGLILLTILFGLSLLLVFPVILYFESTFRCILLRWALLGFATVGT 498
Db 615 VLYRDPVVKSSRELCLVILAGIFLGVCFPTLIATPTTSCYLQRLVLGLSSAMCYSA 674
Qy 499 VTLKLHRLVKV-----FLSRTAQRIPYMTGGRRVMRMLAVILLVFWFELGW 544
Db 675 LVTKTNRIARILAGSKKICKTRKPRFMSAWAQVIAISLVQLTVLVTLII----- 726
Qy 545 TSSVCQNLKLEKISLIGQKTSDDLIFNNCLIDRWDMYTAFA-EFLFLLWGVLYCAYRTV 603
Db 727 -----MEPPMILSYPSIKE--VYLICNTSLNGVAPGVYNGLLIMSCITYAFKTRNV 777
Qy 604 PSATHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHLTVTVTIGLL 663
Db 778 PANENAKYIAFTWYTCILWLAFVIYF-----GSNYKIITTCFAVLSVYVALGCMF 831
Qy 664 IPK-----FSHSSNNPRDDIATEAYEDELDMG-----RSGSYLNSSINSAMSEHSLDP 711
Db 832 TPKMYIIIAKPERNVSFAFT---SDVVRMHVGDGKLPKCRSNTLN----- 874
Qy 712 EDIRDELKLYAQLEIYKRRKMTNNPHLQKRCS-----KGLGRSIRRRITEPTE 764
Db 875 -----IPRKKPGAGNANSKSVSWSEPGGRQAPKQGHVWQRLSVHVKT 919
Qy 765 VSROCKEDKEGADHGTAAGTAKGTALIRKNPPSSNGTGRKSEETLKNRVFSLKSHSTYDVH 824
Db 920 NETACNQ-----TAVIK---PLTKSYQGSCKSLTFSD----- 948
Qy 825 RDQTESSSLPESQEBETTENSTLESLSGKLTQKLEDSEAESESTESVPLVCKSASAHN 884
Db 949 -----ASTKLYNVEEDNTPSAHFSPSPSPSVVHR 980
Qy 885 LSSEKKTCHPRTSMLOKSLSVIASAKETLGLAGTKTAGVEERTKSKOKPLPKOKETNRN 944
Db 981 RGPVATTPPLPHL-----TAEETPLFLADSVTPKGL-----PPLPQQO----- 1021
Qy 945 HNSDNTETKDPAPQNSNPAAEPKPKQKIMKQVRNPTTANSDL-----NPGTTO 996
Db 1022 -----PQQPPQPPQPPK--SLMDQLQGVVTFNGSGIPDFHAVLAGPCTP- 1065
Qy 997 MKDNFDIGEVCPWEVYDLTPGPVPSE--SKVKHVSIVASEMEKNPTFSLKEKSHHKPKA 1054
Db 1066 -----GN-----SLRSLYPPPPPPQHLMPLHLSTFQEEISIPPGEDIDDD- 1108
Qy 1055 AEVCQSNQKRIKAEVCLWESQ-----QSTLEDEKLLISKTPVLPRAKEENGQCPRAAN 1111
Db 1109 -----ERFKLLQBFYVEREGNTTEDELEEEEDLPTASKLTPEDSPALTPPSPRDS 1159
Qy 1112 VCAGQSEELPP--KAVASKTENENLNOIGHQEKKTSSS 1147
Db 1160 VASGSSVSPSPSVESVLCPTPNVTYASVILRDYKQSSS 1197
RESULT 2
MGR5_RAT
ID MGR5_RAT STANDARD; PRT; 1203 AA.
AC P31424;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.
GN GRM5 OR MGLUR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
MEDLINE=92317054; PubMed=1320017;
Abe T., Sugihara H., Nawa H., Shigemoto R., Mizuno N., Nakanishi S.;
"Molecular characterization of a novel metabotropic glutamate
receptor mGluR5 coupled to inositol phosphate/Ca2+ signal
transduction.";
J. Biol. Chem. 267:13361-13368(1992).
[2]
SEQUENCE OF 859-923 FROM N.A., AND ALTERNATIVE SPLICING.
TISSUE=Brain;
MEDLINE=93343913; PubMed=7688218;
Minakami R., Katsuki F., Sugiyama H.;
"A variant of metabotropic glutamate receptor subtype 5: an
evolutionally conserved insertion with no termination codon.";
Biochem. Biophys. Res. Commun. 194:622-627(1993).
-1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CALCIUM SECOND MESSENGER SYSTEM AND GENERATES A CALCIUM-ACTIVATED
CHLORIDE CURRENT.
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 5A (SHOWN HERE) AND 5B; ARE
PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER BY AN INSERTION OF
32 RESIDUES.
-1- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN NEURONAL CELLS OF THE
CENTRAL NERVOUS SYSTEM.
-1- MISCELLANEOUS: ACTIVATED BY QUISQUALATE > GLUTAMATE > IBOTENATE >
TRANS-1- AMINOCYCLOPENTYL-1,3-DICARBOXYLATE.
-1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
STRONGEST, TO MGLUR1.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; D10891; -; NOT_ANNOTATED_CDS.
EMBL; S64315; AAB27666.1; -;
PIR; A42916; A42916.
GCRDB; GCR_0444; -;
GCRDB; GCR_0760; -;
InterPro: IPR000162; -;
InterPro: IPR000202; -;
InterPro: IPR000337; -;
InterPro: IPR001828; -;
Pfam; PF00003; 7tm_3; 1.
PIR; PF01094; ANF_receptor; 1.
PRINTS; PR00248; GPCRMR.
PRINTS; PR00593; MTABOTROPICR.
PRINTS; PR01055; MTABOTROPIC5R.
PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
PROSITE; PS00982; G_PROTEIN_RECEP_F3_4; 1.
G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
Multigene family; Alternative splicing.
SIGNAL 1 20 POTENTIAL.
FT CHAIN 1 20
FT DOMAIN 21 1203 METABOTROPIC GLUTAMATE RECEPTOR 5.
FT DOMAIN 22 578 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 579 601 I (POTENTIAL).
FT DOMAIN 602 615 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 616 636 II (POTENTIAL).
FT DOMAIN 637 647 III (POTENTIAL).
FT TRANSMEM 648 666 IV (POTENTIAL).
FT DOMAIN 667 692 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 693 713 V (POTENTIAL).
FT DOMAIN 714 736 V (POTENTIAL).
FT TRANSMEM 737 758 V (POTENTIAL).
FT DOMAIN 759 771 CYTOPLASMIC (POTENTIAL).

FT	TRANSMEM	772	794	VI (POTENTIAL).
FT	DOMAIN	795	800	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	801	826	VII (POTENTIAL).
FT	DOMAIN	827	1203	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	88	88	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	377	377	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	381	381	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	444	444	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	733	733	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPIC	876	907	MISSING (IN ISOFORM 5A).
SQ	SEQUENCE	1203	AA;	131885 MW; 99CA51E9E11C1EA4 CRC64;

Query Match 3.3%; Score 211; DB 1; Length 1203;
Best Local Similarity 18.7%; Pred. No. 0.0012;
Matches 179; Conservative 128; Mismatches 400; Indels 248; Gaps 36;

QY	219	ETWFHGLRRKWRPHLHRRGPNQGRGLGHSWR-----RKDGLGDKSHKFWPPYLECE	273
DB	332	DVKWFDDYLLKLRPEITNLRNP-----WFQEFQHRFQCRLEGFAQENSKYKNT-----CN	381
QY	274	NG-----SYKPGWLIV--TLSSAIYGL--OPNLVPEFRGV-----MKVDINLQ	311
DB	382	SSLTLRTHVQDSKMGFVINAIYTMAYGLHNMQMSUCPGYAGLCDAMKPDIGRKLDSLM	441
QY	312	KVIDOCSSDGSFSGTHKCHLNSECMPKGLGFLVGLGAYECICKAGFYHPGVLPV-----	366
DB	442	KTNFTGVSGDMILFDENGSDSPGRYEIMNFKEG--KDYFDYINVGSMDNGELKMDDEV	498
QY	367	----NFRRRGPDQHI--SGSTKDVSEE-----AVCLPCREG-----CPFC-----A	403
DB	499	WSKKNNTIIRSCPECKGQIKVIRKEVSCCWTCTCKENEXYVDFDYTKACOLGSWPT	558
QY	404	DD--SPCFVQEDKYLR-----LAISFGQLCMLLDFVSMVWVHFRKAKSIRASGLILLE	456
DB	559	DDITGCDLIPVQDPPPIAAVFPACUGLLATLEVTVIYIIRTPVPPVSSSRRELCY	618
QY	457	TILFGSLLLYFPVVIIVFEPSTFCILLRWARLIGFATVYGTVTIKLHRLKV-----	509
DB	619	IILAGICLGYLCTFCLIAKPKQIYCYLORIGLSPAMSYSALVTKTNRILARILAGSKK	678
QY	510	-----FLSRTAQRIPYTGGRVMRLAVILLVWFVLGWTSSVCONLEKQISL-----	558
DB	679	ICTKKPRFMSACAQ-----LVIAPIIL-----CIQLGIIVALEIME	714
QY	559	---IGOKTSDHLIFNNCLIDRWDTAVA-EFLFLMGVLYCYAVRTVPSAFHEPYMA	614
DB	715	PPDLMHDYPSIREVYLICNTNLGVVTPGLYNGLLILISCTFYAFKTRNVPANFNEAKYIA	774
QY	615	VAVHNELIISAIFHTIRFVLASRLQSDWMLMFLYFAHTHLTVTITGLLLIPK-----FSHS	670
DB	775	FTMTYTCIIWLAFPVIF-----GSNYKIITWCFVSLSATVALGOMFPKVYIILAKP	828
QY	671	SNNPRDDIATE-AYEDELDMGRSGYL--NSSINSANSEHSLEDPEITRDELKIL---YAQ	724
DB	829	ERNVRSATFTSTVVRMHVGDGKSSAASRSSSLNLMKRRGSSGETLRYKDRRLAOKSE	888
QY	725	LEIYKRKKMTITNPHLQKCKSKKL-----GRSIMRRIT-----EIP-ETVSR	767
DB	889	IECTPKSGMNGRATMSSSSNGKSVTWAONEKSTRQHLWRLQWLVHINKENPNQAVI	948
QY	768	QCSKEDEKADHGTAKG-----TALIRKNPPSSSGNTGSKSEETLKNRVF	812
DB	949	KFPKSTENRGPCAAAGGGSGPGVAGAGNAGCTATGGPEPPDAGPKA-----	995
QY	813	SLKKSHTYDHRDQDTESSSLPTESQEEETENSTLESLSGKKLQKLEDSEASTES	872
DB	996	-----LYDVAEAEESFPAAARPSRPSITSLSHLAG-----SAGRTDDDD	1034
QY	873	VPLVCKSASAHNLSSEKKTGHPRTSMLOKLSIVTASAKETLGLAG-----	918
DB	1035	APSLHSETAARSSSSQGLMEQISSVVVTRFTANISLNSMMLSTAAATPGPGTICSSYL	1094

QY 919 -----KTQAGVEERTKSKQPLPKDKETNRNHSNDTETKDPAPNSNPAAEE-PRK 969
Db 1095 IPKEIQLEPTMTTFAE-----IQPLPAIEVT-----GGAQAGTGVSPAQETPTGAESAPCK 1145
QY 970 POKSGIMKQORVNPNTANSNDLPGTQMKDNFDICEVCPWEVYDITPGVPVPSK 1024
Db 1146 PDLEELVALTPPSP-----FRDSDVSGSTTTPNSPVSEALCIPSPK 1187

RESULT 3
NP14_RAT STANDARD: PRT: 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEOLAR PHOSPHOPROTEIN P130 (NUCLEOLAR 130 KDA PROTEIN) (140 KDA
DE NUCLEOLAR PHOSPHOPROTEIN) (NOPP140).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.

TISSUE=Liver;
RX MEDLINE=92323542; PubMed=1623516;
RA Meier U.T., Blobel G.;
RT "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL Cell 70:127-138(1992).
CC -!- FUNCTION: RELATED TO NUCLEOGENESIS, MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH DKC1/NAP57.
CC -!- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN
CC NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE
CC FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOLAR PLASM TO
CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.
CC -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOPP140 IS ONE OF THE
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M94287; AAA41718.1; -;
EMBL; M94288; AAA41719.1; -;
KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
FT DOMAIN 84 570 11 X 12 AA APPROXIMATE REPEATS OF AN

FT REPEAT 84 95 ACIDIC SERINE CLUSTER.
FT REPEAT 127 138 ACIDIC SERINE CLUSTER 1.
FT REPEAT 170 181 ACIDIC SERINE CLUSTER 2.
FT REPEAT 231 242 ACIDIC SERINE CLUSTER 3.
FT REPEAT 274 285 ACIDIC SERINE CLUSTER 4.
FT REPEAT 335 346 ACIDIC SERINE CLUSTER 5.
FT REPEAT 373 384 ACIDIC SERINE CLUSTER 6.
FT REPEAT 434 445 ACIDIC SERINE CLUSTER 7.
FT REPEAT 479 490 ACIDIC SERINE CLUSTER 8.
FT REPEAT 524 535 ACIDIC SERINE CLUSTER 9.
FT REPEAT 559 570 ACIDIC SERINE CLUSTER 10.
FT MOD_RES 567 567 PHOSPHORYLATION (BY CK2).
FT VARIANT 150 150 MISSING (IN NOPP140B).
SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;

Query Match 3.1%; Score 199; DB 1; Length 704;
Best Local Similarity 21.3%; Pred. NO. 0.0026;
Matches 126; Conservative 92; Mismatches 240; Indels 134; Gaps 24;
QY 665 PKFSHSSNNP-----RDIATEAYED---ELDMGR-----SGSYLN 597
Db 64 PKVKLOSNGPVAKKAKKETSSDSSDSEEDKAGVPTQKAAAPAKRASLPQHGAKAAA 123
QY 698 SSINSAMSEHSLDPEDIRDELKLYAQLEIYKRKKMTNNPHLOKKRCKGLGHSIMRR 757
Db 124 KASESSSESESEEEKKKKPVQOKAVKPOAKAVRPPP---KKAPESESESSSE- 179
QY 738 ITEIPETVSROCKEDKEGADHGTAKGTALIRKNPESPESGNTGKSKETLKNRVSLKKS 817
Db 180 -DEAPQT-----QKPKAAATAAKAPTAKQTKAPKPGPPAKPAKAAANGKACGSSSS 230
QY 818 HSTYDHRVDQTEESSSLPTEESQEBETTENTLESLSGKK-LTOKLKDESEARSTESVPLV 876
Db 231 -----SSSSSSDDSEEEKKAAAPLKKTPAKKKVYVAKAPVKVTAAPTQ----- 272
QY 877 CKSASAHNLSE-----KKTGHPRTSMLQKSLSVIASAKEKTLGLAGKTQTAGVEE 927
Db 273 -KSSSESSSESEEEQKKPMKKKAGPYSSVPPPSVSL-----SKKSVG----- 315
QY 928 RTKSOKPLPKDKNRHNNSDNTETKDPAPNSNPAAEPRKPKQSGIMK-QORVNPPTA 986
Db 316 ---AQSPKAAATOPADSSADSSSESD-----SSSEEEKTKPAKTVPVSKTPAKPAPVK 367
QY 987 NSDLNPGTITQMKDNFDICEVCPWEVYDITPGVPVPSKQKHVSIIVASEMEKNPTFSIKE 1046
Db 368 KAE-----SSSDSSDSSDEAPAKPVSAKPSLSPKPAVTPKPPAAKAVATPKQPA----- 419
QY 1047 KSHHKPAAEVCQSNOKRIDKAEVCLWESOGOSTLEDEKLLISK-TPVLPERAKEENG 1105
Db 420 GSGQKPSQKADSSSESESSSESE---EATYKSVTTTPKARVTAKAASLPAKQAPRAGG 476
QY 1106 QPRAANVCAGSEELPPKAVASK-----TENENLNQIGHQEKTKSSSEENVRGYSN 1156
Db 477 DSSSDSSSESEEEKTKPPKPAKKAAGAAVPPKTPVKKAAAESESSSSSSSE-----D 530
QY 1157 SSNFOQPLTSRAEVCVPWEFETPAQPNAGRSVALPASPALSANKIAGPRKEE 1208
Db 531 SSEEKKPKPSKA-----TP-KPQAGKANGVPASQ-----NGKAGRESEE 569

RESULT 4
MGR5_HUMAN STANDARD: PRT: 1212 AA.
AC P41594;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.
GN GRM5 OR MGLUR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RX MEDLINE=94197696; PubMed=7908515;
RA Minakami R., Katsuki F., Yamamoto T., Nakamura K., Sugiyama H.;
RT "Molecular cloning and the functional expression of two isoforms of
RL human metabotropic glutamate receptor subtype 5.";
RL Biochem. Biophys. Res. Commun. 199;1136-1143(1994).
RN [2]
RP REVISIONS..
RA Katsuki F.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 860-952 FROM N.A.
RC TISSUE=Brain;


```
QY 1106 QPRAANVCAG-----QSEELPPKAVASKTENENLQIGHOEKKTSSSEENVNRCGSYSSNNF 1161
DB 1128 OPAAGAAQAACDAARESPAAGPAAAKPDLELVAL-----TPSPFRDSVDSGST- 1178
QY 1162 QQPLTJSRAEVCWPFETPAQPNAGRSVALPAS 1193
DB 1179 -----TPNSPVSEALCIPSS 1194

RESULT 5
MGR1_HUMAN
ID MGR1_HUMAN STANDARD: PRT: 1194 AA.
AC Q13255: Q13255; Q14757; Q14758;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.
GN GRM1 OR MGLUR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE-96029774; PubMed-7476890;
RA Desai M.A., Burnett J.P., Wayne N.G., Schoepp D.D.;
RT "Cloning and expression of a human metabotropic glutamate receptor 1
RT alpha: enhanced coupling on co-transfection with a glutamate
RT transporter";
RL Mol. Pharmacol. 48:648-657(1995).
[2]
SEQUENCE FROM N.A.
MEDLINE-97231349; PubMed-9076744;
RA Stephan D., Bon C., Holzwarth J.A., Galvan M., Pruss R.M.;
RT "Human metabotropic glutamate receptor 1: mRNA distribution,
RT chromosome localization and functional expression of two splice
RT variants.";
RL Neuropharmacology 35:1649-1660(1996).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN
CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR5.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL: U31215; AAA87843.1; -
DR EMBL: U31216; AAA87844.1; -
DR EMBL: L76627; AAB05337.1; -
DR EMBL: L76631; AAB05338.1; -
DR MIM: 604473; -
DR GCRDb; GCR_1825; -
DR GCRDb; GCR_1826; -
DR GCRDb; GCR_1982; -
DR GCRDb; GCR_1983; -
DR InterPro; IPR000162; -
DR InterPro; IPR000337; -
DR InterPro; IPR001256; -
DR InterPro; IPR001828; -
DR Pfam; PF000003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
```

```
DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01051; MTABOTROPICR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS0259; G_PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family; Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 1194 METABOTROPIC GLUTAMATE RECEPTOR 1.
FT DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 593 615 I (POTENTIAL).
FT DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 630 650 II (POTENTIAL).
FT DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 662 680 III (POTENTIAL).
FT DOMAIN 681 706 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 707 727 IV (POTENTIAL).
FT DOMAIN 728 750 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 751 772 V (POTENTIAL).
FT DOMAIN 773 785 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 786 808 VI (POTENTIAL).
FT DOMAIN 809 814 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 815 840 VII (POTENTIAL).
FT DOMAIN 841 1194 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1014 1035 GLN/PRO-RICH.
FT DOMAIN 1067 1081 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1095 1130 SER-RICH.
FT DOMAIN 1142 1194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 887 906 NSNGKSVSWSPGGGVKPG -> KKQPEFSPSQCSA
FT VARSPLIC 907 1194 AQL (IN ISOFORM BETA).
FT VARSPLIC 907 1194 MISSING (IN ISOFORM BETA).
FT VARSPLIC 887 906 NSNGKSVSWSPGGGVKPG -> KKQPEFSPSQCSA
FT CONFLICT 593 593 VQL (IN ISOFORM BETA FROM REF. 2).
FT SEQUENCE 1194 AA; 132376 MW; 970E51AF40584F40 CRC64;

Query Match 3.0%; Score 193; DB 1; Length 1194;
Best Local Similarity 18.7%; Pred. No. 0.01;
Matches 165; Conservative 114; Mismatches 326; Indels 276; Gaps 29;

QY 357 GFYHPGVLPVNNFRRPGDPQHSSTKDYSE-----AYVCLPCREG- 398
DB 498 GTWHEGVUNIDYKIQ---MNKSGVVRVSVCEPCLKQIKVIRKGEVSCCWTACKENE 554
QY 399 -----ADSPCFVQEDKYLR-----LATISFGCLMLDFVSMLVV 438
DB 555 YVDEFTCKACDLGWPNADLTGCEPIPVYLEWSNIEPIAIEAFCIGILVTLFVLIF 614
QY 439 YHFRKAKSIRASGLILLLETLFGSLLYFPVVLVYFEPSTFCILLRWALIGFATVYGT 498
DB 615 VLYRDPVVKSSRELVCILLAGIFLVGVCPETLAKPTTSCYLQRLVLGLSSAMCYSA 674
QY 499 VTLKLRHLVKV-----FLSRTAQRIPTMGTGRVWRMLAVILLVWFELGW 544
DB 675 LVTKTNRIARILAGSKKICKTRKPRFMSAWAQVITIASILISVQLTLVVTLLI- 726
QY 545 TSSVCQNLKQISLIGQCKTSDHLIFNCLDRDYMATAVEF-----LFLMGVYLCVAV 600
DB 727 -----NEPPMPIL-----SYPSIKEVYLICNTSNLGVAPLGYNGLLISCTYAKET 774
QY 601 RTVPSAFHEPRYMAVAVINELIISAIFHTIRFVLASRLQSDMMLLYFAHTHLTVTVTIG 660
DB 775 RNVPANENEAKYIAFTMTTCIIWLAFVPIVF-----GSNYKIITTCFAVSLSVTVALG 828
QY 661 LLLTPK----FSSNSNPRDDIATEAYEDELDMG-----RSGSYLNSINSAWSHS 708
DB 829 CMFTPKMYIIITAKPERNVNSAFTT---SDVVRMHVGDGKLPKCRSNTFLN----- 874
```


QY 525 R-----VWRMLAV-ILLVWFVFLIGHTSSVQONLEKQISLIGOCKTSD-----HLIFNMCL 574
Db 737 SQVVIPTSLIAIQTLMIMVWVPPGTRFYPPDRR-EVILCKIQDMSFLFQSLYNMIL 795
QY 575 IDRWDMYTAFAELFLWGYLYCYAVRT--VPSAFHEPRYMAVAVINELIISAIFITIRF 632
Db 796 IT-----ICTYIAIKTRKIPENFENESKEIFGTYMTTCIIMLAFAPIYF 838

RESULT 7

SIR4_YEAST STANDARD; PRT; 1358 AA.
AC P11978;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE REGULATORY PROTEIN SIR4 (SILENT INFORMATION REGULATOR 4).
GN SIR4 OR STE9 OR ASD1 OR UTH2 OR YDR227W OR YD9934.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=88142836; PubMed=3325825;
RA Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT "Functional domains of SIR4, a gene required for position effect
regulation in Saccharomyces cerevisiae";
RL Mol. Cell. Biol. 7:4441-4452(1987).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=95192063; PubMed=7885847;
RA Davies C.J., Hutchison C.A. III;
RT "Insertion site specificity of the transposon Tn3";
RL Nucleic Acids Res. 23:507-514(1995).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE PROTEINS SIR1 THROUGH SIR4 CONTROL THE EXPRESSION OF
MATING TYPE GENES RESIDENT AT LOCI OF EITHER END OF CHROMOSOME
III. SIR3 AND SIR4 ASSOCIATE WITH THE C-TERMINUS OF RAP1 TO FORM A
DNA-BINDING COMPLEX THAT INITIATES THE REPRESSION AT THE HM LOCI
AND TELOMERES.
CC -1- SUBUNIT: INTERACTS WITH RAP1 C-TERMINUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M37249; AAA20881.1; -;
DR EMBL; U13239; AAC33144.1; -;
DR EMBL; Z48612; CAA88507.1; -;
DR PIR; A29360; A29360.
DR SGB; S0002635; SIR4.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Coiled coil.
FT DOMAIN 1277 1347 COILED COIL (POTENTIAL).
FT VARIANT 994 994 P -> L.
SQ SEQUENCE 1358 AA; 152061 MW; 9C698765964F094E CRC64;

Query Match 2.8%; Score 181; DB 1; Length 1358;
Best Local Similarity 20.8%; Pred. No. 0.051;
Matches 122; Conservative 78; Mismatches 233; Indels 154; Gaps 22;

QY 648 FAHTLTVTVTIGLLLPKFSHS-----SNNPRDDIATE-AYEDELDMGRSGSYLNSSIN 701

Db 94 FPHSNLEKISNKLILLRSKTSAGRIESNPNPNSHDASRLASFQPAFSRHAQOQTSTEN 153
QY 702 S-----ANSEHSLDPEDIRDELKLYAOLEIYKRRKMTNNPHLOK 742
Db 154 SRPVRTIVPISQTNNSFLSGVSKLLSEKIRDSYKEILG-----INLANEQPVLEK 206
QY 743 KRCSKKG---LGRSIMRRITE---IPETVSROCSKDEKADHGHTAGKTALIRKNPPES 796
Db 207 P--LKKGSADICASYISLTKDSIRKDTVEEK--KEEKLNIKNFAHSDSL--SVPKVA 260
QY 797 GNTGSKKEETLNKRVFSLKSHSTYDHDQTEESSESSLPTESEEEETTSTLESLSGKK 856
Db 261 GDSGISPEES--KARSPGIAPNAIQTEVYGINEESTNERLEINQEKPV----- 307
QY 857 LQKLEKDESEAEESTVPLVCK--SASAHLNLSSEKKTGHPRTSMLOKLSLVSASAKEKTL 914
Db 308 ---KLDENS--ANSTVASALDTNGTSATTETLSKKIVSPKKAVIDO---DKITLHDEKTL 361
QY 915 GLAGKTQTAGVEERTKSQKPL-----PKDKETNRNHSNSDNTETK-- 954
Db 362 A-PSKHQIPITSEKMKEDADLARMELIKSPHLSKPADRPQGRNRSRNFSTDEETTKLA 420
QY 955 ---DPAPQNSNPAEERPKQSGIMKQQRVNTTANSDLNPGTTOMKDNFDIGEVCPWEV 1011
Db 421 FLVEYEGEENYNSTSRSTEKKNDMNTSAKNNGENKKG-- 460
QY 1012 YDLTGPCVPSESKVQKHVSIVASEMEKNPTFSLKEKSHHKPKAAEVCQOSNOKRIDKAEV 1071
Db 461 -----KRPEIMSTEAHNVKVYTEETTKQIQSVIRDRGRV 494
QY 1072 CLWESQSGOSILEKLLISKTP-----VLPERAKEENGQOPRAANVCAGQ---SEELP 1121
Db 495 -LQKVQGESHIDSRNTLNVTFSKRPQLGEIPNPKKHKPNSEGRTPNISNGTINQKLE 553
QY 1122 PRXAVASKTENENLQIHQEKKTSSSEENVRSYSSNNNFQOPLFSR 1168
Db 554 PK-----EIVRDILHTKESSEAKKTIONPLNKSQNTALPSTHK 592

RESULT 8
MRSP_STA0U STANDARD; PRT; 1637 AA.
AC P80544; Q9ZF62;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METHICILLIN-RESISTANT SURFACE PROTEIN PRECURSOR.
GN PLS.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE 1061;
RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
RT "pls, a large repeat-rich surface protein of methicillin resistant
Staphylococcus aureus";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=ISOLATE 1061;
RX MEDLINE=96270743; PubMed=8665912;
RA Hilden P., Savolainen K., Tynnelae J., Vuontola M., Kuusela P.;
RT "Purification and characterisation of a plasmin-sensitive surface
protein of Staphylococcus aureus";
RL Eur. J. Biochem. 236:904-910(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way


```
QY 1132 ENLNQIGHQEKTSSEENVGRSYNNFQPLTSRAEVCWETETPAQPNAGRS---V 1188
Db 610 -----GKKTEISEKESKADMKHLREKVKSTRKESLQHLNVTAKKPARVSKDVE 661
QY 1189 ALPAS-SALSANKIAGPKREEIWDSF 1213
Db 662 DVPASKAKEGTEDVSPTKQKSPISF 687

RESULT 10
ANK2_HUMAN
ID ANK2_HUMAN STANDARD; PRT: 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANKYRIN 2 (BRAIN ANKYRIN) (ANKYRIN B) (ANKYRIN, NONERYTHROID).
GN ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
RN [2]
SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RP TISSUE=Brain stem;
RX MEDLINE=91302466; PubMed=1830053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT Ankyrins reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).
RN [3]
REVISED.
RA Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene.";
RL Genomics 10:858-866(1991).
CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
CC BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
CC THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC CELLS THROUGHOUT THE BRAIN.
CC -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC AND FUNCTION (POTENTIAL).
CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/)
```

or send an email to license@isb-sib.ch).

```
CC EMBL; Z26634; CAB42644.1; -.
DR EMBL; X56957; CAA40278.1; -.
DR EMBL; X56958; CAA40279.2; -.
DR EMBL; M37123; AAA62828.1; -.
DR PIR; S14533; S14533.
DR PIR; A39643; A39643.
DR PIR; B39643; B39643.
DR PIR; S14569; S14569.
DR HSP; Q00420; IAWC.
DR MIM; 106410; -.
DR InterPro; IPR000488; -.
DR InterPro; IPR000906; -.
DR InterPro; IPR002110; -.
DR Pfam; PF00791; ZU5; 1.
DR Pfam; PF00023; ank; 22.
DR Pfam; PF00531; death; 1.
DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation; Multigene family;
FT REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A (APPROXIMATE).
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A (APPROXIMATE).
FT REPEAT 1892 1902 REPEAT A.
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH DOMAIN.
FT VARSPLIC 1039 1039 Q -> OFLGKHLPTAPPPLNEGSLVSRILQLGPPGTK
FT VARSPLIC 1444 3528 (IN ISOFORM 2).
FT VARSPLIC 475 476 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT CONFLICT 971 971 GQ -> PE (IN REF. 4).
FT CONFLICT 3581 3582 I -> S (IN REF. 2).
FT CONFLICT 3586 3586 QY -> HA (IN REF. 2).
FT CONFLICT 3924 3924 I -> Y (IN REF. 2).
SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;
```

Query Match 2.8%; Score 177.5; DB 1; Length 3924;
Best Local Similarity 21.3%; Pred. No. 0.3;
Matches 123; Conservative 85; Mismatches 230; Indels 139; Gaps 26;

QY 691 RSGSYL--NSSINSAWSEHSL-----DPEDIRDELKLYAQLEI----- 727
DB 1534 RSGCTRDESSVSSRSEGLVEEWIVSDDELEEARQK--APLEITEYPCVEVRIDKE 1591
QY 728 -----YKRRKMITNPP-----HLQK-----KRCSSKKGIGRSIMRRITEPE 763
DB 1592 IKGKVEKDTGLVNYLTDLNTCVPLPKEQLQTVODKAGKKCEALAVGRSSEKGDIPP 1651
QY 764 TVSRQCSKDEKADHCTAGTALIRKNPPSSNGTCKSKEETLKNRV--FSLKKSSTY 821
DB 1652 DETQSTOKQHPSL--GIKKP--VRRKLKEQ-----KQEEGLQSAEAKAELKGSSE 1702
QY 822 DHVRDQTEESSSLPTESQEBETTENTSTLESISK-KLTKRLKEDSEAESVPLVCKS- 879
DB 1703 SLGEDPGLAPEPLTVKATSPLEETPGISIKDKVKALQKRVED-EQKGRSKLPVIRYKGK 1761
QY 880 -----ASAHNLSEKKTGHPTSMLOKLSVSIASAKK---TLGLAGTKTQ-- 921
DB 1762 EDVPKKTTHRPHAPASPSLKSERHAPGSPKTERHSTLSSAKTERHPPVSPSSKTEKH 1821
QY 922 -TAGVEERTKSQPLPKDKETNRNHSNDNTETKDPAPONSNAEPRKPKQSG-IMKQQ 979
DB 1822 SPVSPSAKTERHSPASSSSKTEKHSVPSTKTERHSPVSPSGKTDKRP 1881
QY 980 RVNP---TTANSDLNPTQTKMKNFDIGECVPEVYDLTPG-----PVPSESKVKOHV 1029
DB 1882 PVSFSGRTEKHPVSPORTEKR-----LPVPSGRTDKHQPVSTAGKTEKHL 1928
QY 1030 SIVAS-EMEKNTFTLSKEKSHHAKPAEVCQ-----QSNQKRIDKAEVCLWE----- 1075
DB 1929 PVSFSGRTEKHPVSPORTEKR-----LPVPSGRTDKHQPVSTAGKTEKHL 1928
QY 1076 --SQGQSILEDKLLISKTPVPERAKEENGQPPRAANVAGQSEELPPRAVASKTENEN 1133
DB 1988 POEKGVRVEKE-----KGPILTQREAO-----KTENQTKRGQRLPVTGTAEKRGVR 2036
QY 1134 LNOIGHQEKTSSEENVRGSYNSNNFOOPLTSRAE 1170
DB 2037 VSSIGVKKEDAAGGKEV-----LSHKIPVQSVPE 2068

RESULT 11

QY 11 BOVIN
AC P35662; CYLL_BOVIN STANDARD; PRT; 667 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYLICIN I (MULTIPLE-BAND POLYPEPTIDE 1).
GN CYLL OR CYL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cylicin, a basic protein of
the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC -/- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -/- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -/- TISSUE SPECIFICITY: TESTIS.
CC -/- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z22779; CAAB0456.1; -.
DR PIR; S35913; S35913.
DR PIR; A40713; A40713.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 287 589 9 APPROXIMATE TANDEM REPEATS.
FT REPEAT 287 305 1.
FT REPEAT 306 337 2.
FT REPEAT 338 368 3.
FT REPEAT 369 405 4.
FT REPEAT 406 442 5.
FT REPEAT 443 475 6.
FT REPEAT 476 516 7.
FT REPEAT 517 547 8.
FT REPEAT 548 569 9.
FT DOMAIN 617 667 PRO-RICH.
FT SEQUENCE 667 AA; 74817 MW; CBF66EA462243D91 CRC64;

Query Match 2.8%; Score 177; DB 1; Length 667;
Best Local Similarity 20.2%; Pred. No. 0.033;
Matches 103; Conservative 89; Mismatches 180; Indels 138; Gaps 19;

QY 665 KFSHSSNNPRDDIATEAYEDELDMGRSGSYLNSINSANSEHSLDPPDIR-DEUKKLYAQ 724
DB 228 KYSKSSKNSDAVSETCKNSNVGLM-VHLGES-----DAESMEFDMWLKNYSQ 276
QY 725 LEIYKRRKMITNPHLOKRCCKSK---KGLGRSIMRRITEIPETVSROCSEKDEGADHGT 781
DB 277 -----NNSKPTKKDAKKDAKKGSD-----AESVDSKAKKDKKATKDT 317
QY 782 AKGTALIRKNPPSSNGT---GKSKEETLKNRVFLSK---SHSTYDHRDQTEESSLP 835
DB 318 KKGAKKDTSTDAESGSDSKAKKGKESKKDKKDAKDAASDAESGSDSKAKKKGK 377
QY 836 TESQEE-----ETTENTSTLESLSGKLTQKLEDS-----EAEATESVP 874
DB 378 KDSKKNKKKDAKDAESTDAESGSDSKAKKDKKDKKDKKDKKDKKDAESTDA-- 435
QY 875 LVCKSASAHNLSEKKTGHPTSMLO--KSLSVIASAKEKTLGLAGTKTQTAGVEERTKSQ 933
DB 436 ---ESGDSKNAKKDKKDKKDKKDAKDAVDTDAESEGDAAKSK---KDSKDKK 488
QY 934 PLPKDKETNRNHSNDNTETKDPAPQNSNPAEPRKPKQSGIMKQQRVNPPTANSDLNPG 993
DB 489 DLKKDQKKPKAMSKESTETESDWESKVKRDSKKDTKTA---KKATESSGAESD--- 541
QY 994 TTOMKDNFDIGECVPEVYDLTPGPVPSSEKVKOHVSIASEMEKNTFTLSKEKSHHAKP 1053
DB 542 -----VSSKRYLKTETFKSSDAE-----SEESLEKP- 568
QY 1054 AAEVCQSQSNOKRIDKAEVCLWESQGSQSILEDKIL-LLSKTPVLPVPERAKEENGQ- 1106
DB 569 -----GSKRVEDSATDTSKKDAVEPKRGIMKPSRRTTTFKEGKKKIGTGVRVPSRE 621
QY 1107 -----PRAANVAGQSEELPPK 1123
DB 622 RPPLPCEPILPSPRVKRLCRCQMPPPPPK 651

RESULT 12

QY 12 HUMAN
AC P12270; TPR_HUMAN STANDARD; PRT; 2349 AA.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

01-OCT-2000 (Rel. 40, Last annotation update)
DT NUCLEOPROTEIN TPR.
DE TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "Tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RJ J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE=88262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TPR-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -!- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -!- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X66397; CAA47021.1; -;
CC EMBL; Y00672; CAA68681.1; -;
CC PIR; S00928; S00928.
CC MIM; 189940; -;
CC
CC Heptad repeat pattern: Coiled coil; Proto-oncogene;
KW Chromosomal translocation; Nuclear protein; Transport.
FT DOMAIN 78 360
FT COILED COIL (POTENTIAL).
FT DOMAIN 422 571
FT COILED COIL (POTENTIAL).
FT DOMAIN 575 628
FT COILED COIL (POTENTIAL).
FT DOMAIN 805 855
FT COILED COIL (POTENTIAL).
FT DOMAIN 834 869
FT COILED COIL (POTENTIAL).
FT DOMAIN 934 979
FT COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064
FT COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166
FT COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241
FT COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304
FT COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434
FT COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595
FT COILED COIL (POTENTIAL).
FT DOMAIN 1527 530
FT POLY-SER.
FT DOMAIN 1833 1836
FT POLY-GLU.
FT DOMAIN 1957 1964
FT POLY-ASP.
FT DOMAIN 2295 2298
FT POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;

Query Match 2.8%; Score 176; DB 1; Length 2349;
Best Local Similarity 20.7%; Pred. No. 0.19;
Matches 123; Conservative 99; Mismatches 239; Indels 134; Gaps 24;
Qy 707 HSLDPEDIRDELKLYIAQLEIYKRRKMTNNPH-----LQKRCSSKGLGRSIRMRI 758
Db 1364 HTKRIQQLTEEIGRLKA--EARSNASLTNNQNLQSLKEDLNKVRTEKTIQKDLDAKI 1421
Qy 759 TEIPETVSKQCKEDKADHCT-----AKGTALIRKNPPSSG----- 797
Db 1422 IDIQEKV-KTIQVKKIGRRYKTYEELKAQDKVMETSQAOSGDHQHVSQVEMQELK 1480
Qy 798 ---NTGKSKEETLKNRVFLSKKSHSTYD-----HVRDQT-----EESSLPTSEOEETTEN 846
Db 1481 ETLNQAETKSKSLESOVENLQKLTSEKETEARNLQEQTVQLQSELSRLQDLODRTTQE- 1539
Qy 847 STLESLSKKTKQLKEDSEAESESTESVPLVCKASAH-----NLSSEKTKGHPTSMQL 900
Db 1540 -----EQLRQQITE--KEEKTRKATVAASKTAHLAGVKDQLTKNEELKQRNGALD 1589
Qy 901 K---SLSV-IASAKETLTGLAGKT-----QTAGVEERTKSOKPLPKDKETNRN---- 944
Db 1590 QOKDELVRITALKSQYEGRISRLERELREHOERHLEORDEPOEPSNKVPEQOQOITLKT 1649
Qy 945 -----HSNSD-NTETKDPAPQNSPAE-----EPRKPKQSGIMKQOQRYNP 983
Db 1650 TPASGERGIASTSDPTTANIKPTPVVSTPSKVTAAAMAGNKTPTPRASIRPMVTPATVNP 1709
Qy 984 TTANSLNPGTQTKDNDFDICEVCPWE---VYDLTPGVPVPS-KVQKHS--IVASME 1037
Db 1710 TTTTATVMTPTQESQAMQSEGVEHVPGVFGSTGSGVRSTSPNVQPSISQPILTVOQQ 1769
Qy 1038 KNPFLSKKSHHKPKAAEVCQSQKQKIDRAEVLWESQGSILEDEKLLISKTPVLP 1097
Db 1770 TQATAFVQPTQOSHPIEPANQELSSNIVEVQSSPVERPSTSTAVFTVTSATPSSSLPK 1829
Qy 1098 RAKEENGQGP-RAANVCAGQSEELP-PRKAVAS-----PLTSRAEVCPEFETPAQNAGR 1186
Db 1830 RTREEDSTTEASDSDVDIVEMPLPKKLSVTPGVTEEVMAEESTDGEVETQVYNQD 1889
Qy 1142 KKTSSSENVRGSSNNFNQ-----PLTSRAEVCPEFETPAQNAGR 1186
Db 1890 SQDSIGEGVTQGYTPMEDSETSQSLQIDLGQLSDQQT-----TTSSQDQCGK 1939
RESULT 13
CASR_MOUSE
ID CASR_MOUSE STANDARD; PRT; 1079 AA.
AC Q9QY96; O08968; O88519; Q9QY95; Q9QZ08; Q9R1D6; Q9R1Y2;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
DE CELL CALCIUM-SENSING RECEPTOR).
GN CASR OR GPRC2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6; TISSUE=Kidney;
RX MEDLINE=20092890; PubMed=10625662;
RA Oda Y., Tu C.-L., Chang W., Crumrine D., Koemueves L., Mauro T.,
RA Elias P.M., Bikie D.D.;
RT "The calcium sensing receptor and its alternatively spliced form in
RT murine epidermal differentiation.";
RL J. Biol. Chem. 275:1183-1190(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
RC STRAIN-BLACK SWISS X 129/SVJ; TISSUE=Kidney;
RX MEDLINE=20119279; PubMed=10652312;


```
Db 777 IGYTCLLAIAICFFFAFKSRKLPENNEAKFITFSM---LIFFIVWISFIPAYASTYGV 833
Qy 634 LASRLQSDWMLLYFAHTHLTVTVTIGLLLPKFSSNN---NPRDDIATEAYEDELDM- 689
Db 834 SAVE-----VIALAASFGLLACIFFNKVYIILFKPSRNTIEVRSSTAHAFAKVAARAT 888
Qy 690 -----GRSGSYLNSINSAMSHSLDPEDIRDELKLLVAQLIYKRKKMI 734
Db 889 LRRPNTSRKRSSSLGGSTGPNSSSTSSKSNSEDRFPQPEROKQOQPLATQOEQOQPL 948
Qy 735 TNNPHLOKK---RCSKK---GLG-----RSIMRR-----ITEIPETV 765
Db 949 TLOPQQQQQOPRQKQKQVIFSGVTVTSLSFDEPKQKNAHNRNSLEAQNNDTL 1008
Qy 766 SR-----QCSKDEKADGHTAKGTALIRKNPPSSGNTGSKKEETLKNRVFSLKKS 817
Db 1009 NRHQALLPLQCAEADSEMTIQTGLQGPVGDHPQPEI-----ESPDMSPALVVSTSR 1062

RESULT 14
AC15_MOUSE STANDARD: PRT: 1131 AA.
P35601:
01-JUN-1994 (Rel. 29, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
ACTIVATOR 1 140 KDA SUBUNIT (REPLICATION FACTOR C LARGE SUBUNIT) (A1
DE 140 KDA SUBUNIT) (RE-C 140 KDA SUBUNIT) (ACTIVATOR 1 LARGE SUBUNIT)
DE (A1-P145) (DIFFERENTIATION SPECIFIC ELEMENT BINDING PROTEIN)
DE (ISRE-BINDING PROTEIN).
GN RFC1 OR RECC1 OR ISEF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=94089669; PubMed=8265586;
RA Burdello P.D., Utani A., Pan Z., Yamada Y.;
RT "Cloning of the large subunit of activator 1 (replication factor C)
RT reveals homology with bacterial DNA ligases";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11543-11547(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158835; PubMed=8114700;
RA Luckow B., Bunz F., Stillman B., Lichter P., Schuetz G.;
RT "Cloning, expression, and chromosomal localization of the
RT 140-kilodalton subunit of replication factor C from mice and
humans.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS;
RX MEDLINE=95388065; PubMed=7659092;
RA McGehee Habener J.F.;
RT "Differentiation-specific element binding protein (DSEB) binds to a
RT defined element in the promoter of the angiotensinogen gene required
RT for the irreversible induction of gene expression during
RT differentiation of 3T3-L1 adipoblasts to adipocytes.";
RL Mol. Endocrinol. 9:487-501(1995).
RN [4]
RP SEQUENCE FROM N.A.
RA Haque S.J.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-565 FROM N.A.
RC STRAIN=LAF1;
RA Lössie A.C., Haugen B.H., Wood W.M., Camper S.A., Gordon D.F.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 354-528 FROM N.A.
```

```
RX MEDLINE=90229765; PubMed=1691767;
RA Haque S.J., Kumar A., Fischer T., Rutherford M.N., Williams B.R.;
RT "Evaluation of inter- and intramolecular primary structure homologies
RT of interferons by a Monte Carlo method.";
RL J. Interferon Res. 10:31-31(1990).
CC -!- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE
CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
CC PCNA AND ACTIVATOR 1. THE 140 SUBUNIT BINDS TO THE PRIMER-
CC TEMPLATE JUNCTION.
CC -!- SUBUNIT: HETEROPTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND
CC 36.5 KDA THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ACTIVATOR 1 140 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; U01222; AAA21643.1; -
CC EMBL; X72711; CAA51260.1; -
CC EMBL; U36441; AAA79698.1; -
CC EMBL; U07157; AAC52140.1; -
CC EMBL; U15037; AAB60452.1; -
CC MGI; MGI:97891; Reccl.
CC InterPro; IPR001357; -
CC Pfam; PF00533; BRCT; 1.
CC PROSITE; PS0172; BRCT; 1.
CC DNA replication; ATP-binding; Transcription regulation; DNA-binding;
KW Activator; Nuclear protein.
FT DOMAIN 399 477 BRCT.
FT NP_BIND 635 642 ATP (BY SIMILARITY).
FT ZN_FING 734 751 C2HC-TYPE (POTENTIAL).
FT DOMAIN 1104 1108 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 354 528 INTERFERON-STIMULATED-RESPONSE-ELEMENT
FT BINDING REGION.
FT CONFLICT 66 66 Y -> N (IN REF. 3).
FT CONFLICT 187 187 E -> EPDFCLSLIFGIQ (IN REF. 4).
FT CONFLICT 254 254 V -> A (IN REF. 5).
FT CONFLICT 559 559 N -> S (IN REF. 4).
FT CONFLICT 614 614 MISSING (IN REF. 3 AND 4).
FT CONFLICT 945 945 S -> N (IN REF. 1).
FT CONFLICT 1071 1071 T -> A (IN REF. 3).
FT CONFLICT 1104 1104 K -> Q (IN REF. 4).
SQ SEQUENCE 1131 AA; 125984 MW; A6F4F970A7F9EE94 CRC64;

Query Match 2.7%; Score 174.5; DB 1; Length 1131;
Best Local Similarity 23.6%; Pred. No. 0.087;
Matches 108; Conservative 76; Mismatches 157; Indels 117; Gaps 23;

Qy 761 IPETVSRQCKEDKADGHTAKGTALIRKNPPSSGNTGSKKEETLKNRVFSLKSHST 820
Db 17 VNETHV-----KNEKTKASEGVKGGKGVKEAKVNNSG-----KEDASKPKQHS-KKKRII 65
Qy 821 YDHVRDQTEESSSLPTESQBEETTE-----NSTLESLGK--KLTQKLKEDSEAESTES 872
Db 66 YD-----SDSEETVQVKNNAKKSEKLSLSYKPGKVSQKDPVTVYVSETDE 112
Qy 873 VPLVKCSAHLNLSSEKKTG-----HPRTSMLOKSLSVIASAKETLGLACKGTOTA---- 923
Db 113 DDFVCKKAA-----SKSKENGVTNSYLGTSNVKKNEENV-KTKNKPSPILKTPITSVDLY 167
Qy 924 -GVEERTKSKQPL--PKDKETNRNHSND--NTETKDPAPQNSNPAEPRKPKQSGIMKQQ 979
Db 168 FGTESVQRSGKKMVTSKRKSSQNTDSRLNDEATAKQLQDLDAELERQ-----LHEDE 222
Qy 980 RVNPTTANSDLNPGTTQMKDNFDIGEVCPWEVYDLTFCGPVPSSEKVKQHVIVASEMEKN 1039
```

Db 223 EFARTLALLDEEPKIKKARKDSEGE-----ESFSSVQDDLS--KAEKQKS 266
QY 1040 PT-----FSLKEKS-----VCOOSNOKRI-----DRAEYCLWESQGQSI 1081
Db 267 PNKAELFSTARKTYPKAGHGKGRASEDAKQPCSAURKEACSPKASAKLALMKAREES 326
QY 1082 LEDEKLLISKTPVLPRAKEENGCGPRAANVCAGQSEELPPKRAVAKTENENLNQICHQE 1141
Db 327 YNETELLAAR-----RKESATEPKG-----EKTTPKTKVSPKRESVSPDESEK 371
QY 1142 KTSSESSENVGRSYNSNFQOP-----LTSRAEVC 1172
Db 372 KRTNYQ-----AYSYLNREGPKALGSKPEIPKAENC 403

RESULT 15

NPH_HUMAN STANDARD; PRT; 1020 AA.
P12036;
01-OCT-1989 (Rel. 12, Created)
01-OCT-1989 (Rel. 12, Last sequence update)
01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN NEFH OR NFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328981; PubMed=3138108;
RA Lees J.F., Sheidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT "The structure and organization of the human heavy neurofilament
subunit (NF-H) and the gene encoding it.";
RL EMBO J. 7:1947-1955(1988).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15306; CAA33366.1; -
DR EMBL; X15307; CAA33366.1; JOINED.
DR EMBL; X15308; CAA33366.1; JOINED.
DR EMBL; X15309; CAA33366.1; JOINED.
DR PIR; S00979; QFHUH.
DR MIM; 162230; -
DR InterPro; IPR001664; -
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW Phosphorylation.
FT DOMAIN 1 100 HEAD.
FT DOMAIN 101 413 ROD.
FT DOMAIN 414 1020 TAIL.

FT DOMAIN 101 132 COIL 1A.
ET DOMAIN 133 145 LINKER 1.
FT DOMAIN 146 244 COIL 1B.
FT DOMAIN 245 266 LINKER 12.
FT DOMAIN 267 288 COIL 2A.
FT DOMAIN 289 292 LINKER 2.
FT DOMAIN 293 413 COIL 2B.
SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCF1D4 CRC64;

Query: Match 2.7%; Score 174; DB 1; Length 1020;
Best Local Similarity 20.6%; Pred. No. 0.081;
Matches 101; Conservative 79; Mismatches 229; Indels 82; Gaps 17;

QY 754 IMRRIITEIPETVSRQSKEDKAGDHGTAKGTALIRKNPP-ESSNGTGSKEETLKNRVF 812
Db 466 VTEEVTEEEKEAKEEGEGEGEEAEGGEEETKSPPAEEAASPEAKEKSPVKEEAK 525
QY 813 SLKSHS-----TYDHRDQTESSSLTSESQEEETTE-NTLSLSGKLLT 858
Db 526 SPAEAKSPEKEEAKSPAENVKSPKAKSPAEEAKSPKAKSPEKEEAKSPAENVKSPKAK 585
QY 859 QKLKEDSEAESESTESVPLVCKSAHANLSSEKKTGHPRTSMLOKSLSVIA--SAKETLGL 916
Db 586 SPAKEAKSPAEEAKSPEKAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPTKEE 645
QY 917 AGKTOTAGVEER-----TKSQKPLPKDKETNRNHSNDNTETKDPAPQNSNPAEPRKP 970
Db 546 AKSPEKAKSPEKEEAKSPEKAKSPEKAKSPEKAKSPEKAKSPEKAKSPEKAKSPEKAKSP 705
QY 971 QKS-GIMKQORVNPTTANSDLNPGTTOMKDNFDIGEVCPEWVYDLTPGVPVPSKVKOKHV 1029
Db 706 ERAKSPVKEEAKSPEKAKS-----PVKEEAKTPKA 736
QY 1030 SIVASEMEKNPTFSLKEKSHHKPKAAEVCOOSNQKRIDRAEVCLEWESOGSILEDEKLLI 1089
Db 737 KSPVKEEAKSPE-----RAKSPKAKTLDV--KSPEAKTPAKE-----EARSADKPEK--- 784
QY 1090 SKTPVLPRAKEENGCGPRAANVCAGQSEELPPKAVASK--TENENLNQIGHQEKTTSSS 1147
Db 785 AKSPVKEEAKSPEKAKSPLKADAKAPE-KEIPKKEEVKSPVKEEKPQEVKVKPEPKKAE 843
QY 1148 EENVGRSYNSNNFQQPLTSRAEVCPEWETP---AQPNAGRSVLPASSALSANKIAGP 1204
Db 844 EEKAPATPKT-----EEKKDSKKEEAP-KKEAPKPVKEEKPEAVEKPKESKVEA----- 892
QY 1205 RKEEIWDSFKV 1215
Db 893 KKEEAEEDKKKV 903

Search completed: November 1, 2001, 10:20:45
Job time: 378 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:13:32 ; Search time 61.87 Seconds
(without alignments)
2598.200 Million cell updates/sec

Title: US-09-775-181-2
Perfect score: 6382
Sequence: 1 MGAWAYPELLCLLQLGLG.....LSANKIAPRKKEIWDSEFKV 1215

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp-invertebrate.*
- 6: sp-mammal.*
- 7: sp-mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-unclassified.*
- 13: sp-vertebrate.*
- 14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2811	44.0	544	Q9ULT3	Q9ult3 homo sapien
2	1532.5	24.0	317	Q9GKT3	Q9gkt3 macaca fasc
3	571.5	9.0	669	Q9VR40	Q9vr40 drosophila
4	307	4.8	1677	Q9VKA3	Q9vka3 drosophila
5	281	4.4	264	Q9WLP3	Q9wlp3 drosophila
6	279.5	4.4	433	Q45500	Q45500 caenorhabdi
7	254	4.0	1221	Q9Y133	Q9y133 drosophila
8	238.5	3.7	176	Q9WIP4	Q9wip4 drosophila
9	227.5	3.6	1305	Q9WPS7	Q9wps7 drosophila
10	225	3.5	5327	Q76891	Q76891 drosophila
11	210.5	3.3	1199	Q9EPV6	Q9epv6 mus musculus
12	209	3.3	1160	Q9ESC8	Q9esc8 mus musculus
13	205.5	3.2	3111	Q9VH10	Q9vh10 drosophila
14	205	3.2	1183	Q9W3D1	Q9w3d1 drosophila
15	204.5	3.2	6632	Q17362	Q17362 caenorhabdi
16	204.5	3.2	6642	Q01761	Q01761 caenorhabdi
17	201.5	3.2	700	P82179	P82179 canis famil
18	201.5	3.2	1200	Q25388	Q25388 loligo peal
19	200	3.1	2081	Q9LH98	Q9lh98 arabidopsis

20	198.5	3.1	1812	5	Q9VZD9	Q9vzd9 drosophila
21	198	3.1	877	4	Q9UGT0	Q9ugt0 homo sapien
22	198	3.1	6677	5	Q9N435	Q9n435 caenorhabdi
23	195	3.1	1338	11	Q9JIX8	Q9jix8 mus musculus
24	194.5	3.0	3488	5	P91257	P91257 caenorhabdi
25	193	3.0	1481	5	Q9VFR6	Q9vfr6 drosophila
26	190	3.0	1390	5	O77033	O77033 dictyosteli
27	189	3.0	808	4	Q9UK88	Q9uk88 homo sapien
28	187	2.9	1046	5	P91176	P91176 caenorhabdi
29	186	2.9	522	10	Q9FXB5	Q9fxb5 arabidopsis
30	185	2.9	3201	5	Q9W0U2	Q9w0u2 drosophila
31	183	2.9	852	10	Q9SZ55	Q9sz55 arabidopsis
32	183	2.9	976	5	Q9V485	Q9v485 drosophila
33	182.5	2.9	1163	4	Q9UHB7	Q9uhb7 homo sapien
34	181.5	2.8	1698	2	Q9LC00	Q9lc00 staphylococ
35	181	2.8	2151	5	Q9Y1L3	Q9y1l3 drosophila
36	181	2.8	2151	5	Q9VPL6	Q9vpl6 drosophila
37	181	2.8	2151	5	Q9NG79	Q9ng79 trichomonas
38	181	2.8	5322	5	Q9NI64	Q9ni64 drosophila
39	179.5	2.8	423	5	Q9VVH1	Q9vvh1 drosophila
40	179.5	2.8	1230	5	Q20626	Q20626 caenorhabdi
41	179.5	2.8	3261	4	Q9Y556	Q9y556 homo sapien
42	179	2.8	2031	13	Q9PVP4	Q9pvp4 xenopus lae
43	178.5	2.8	1020	5	Q27104	Q27104 trichomonas
44	178.5	2.8	1151	5	Q9VAY4	Q9vay4 drosophila
45	178	2.8	2451	5	Q9VG05	Q9vg05 drosophila

ALIGNMENTS

RESULT 1

ID Q9ULT3 PRELIMINARY; PRT; 544 AA.
AC Q9ULT3;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE KIAA1136 PROTEIN (FRAGMENT).
GN KIAA1136.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC [ISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the GeneMark analysis
from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032962; BAA86450.1; -.
FT NON_TER 1
SQ SEQUENCE 544 AA; 60195 MW; DA1175E7926041F7 CRC64;

Query Match 44.0%; Score 2811; DB 4; Length 544;

Best Local Similarity 99.8%; Pred. No. 8.4e-196;

Matches 543; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 672 NNPDDIATEAYEDELDMGRSGSYLNSSINSANSEHSLDPEDIRDELKLYAQLEYKRR 731

DB 1 NNPDDIATEAYEDELDMGRSGSYLNSSINSANSEHSLDPEDIRDELKLYAQLEYKRR 60

QY 732 KMITNPHLOKKCKSKGLGRSTMRITETIPETVSRQCSKEDKEDAGHTAGTALIRKN 791

DB 1 KMITNPHLOKKCKSKGLGRSTMRITETIPETVSRQCSKEDKEDAGHTAGTALIRKN 120

QY 792 PPSSNGTGTGSKETLKNRVFSLKSHSTYDHRDTERSSSLPTQSEETTENSTLES 851

DB 121 PPSSNGTGTGSKETLKNRVFSLKSHSTYDHRDTERSSSLPTQSEETTENSTLES 180

```
Qy 852 LSGKLTOKLKEDSEAEESTESVPLVCKSASAHNLSSSEKKTGHPRTSMLQKLSLVIASAKE 911
|||||
Db 181 LSGKLTOKLKEDSEAEESTESVPLVCKSASAHNLSSSEKKTGHPRTSMLQKLSLVIASAKE 240
|||||
Qy 912 KTLGLAGKTQTAGVEERTKSKPLPKDKETNRHNSDNTETKDPAPQNSNPAEPRKPQ 971
|||||
Db 241 KTLGLAGKTQTAGVEERTKSKPLPKDKETNRHNSDNTETKDPAPQNSNPAEPRKPQ 300
|||||
Qy 972 KSGIMKQORVNPPTTANSNLNPGTTQMKDNFDIGEYCPWEVYDLTPGVPSESKVKHYSI 1031
|||||
Db 301 KSGIMKQORVNPPTTANSNLNPGTTQMKDNFDIGEYCPWEVYDLTPGVPSESKVKHYSI 360
|||||
Qy 1032 VASEMEKNPTSLSEKSHHKPKAAEVCOOSQKRIKAEVCLWESQOSILEDEKLLISK 1091
|||||
Db 361 VASEMEKNPTSLSEKSHHKPKAAEVCOOSQKRIKAEVCLWESQOSILEDEKLLISK 420
|||||
Qy 1092 TPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOIGHOEKKTSSSENV 1151
|||||
Db 421 TPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOIGHOEKKTSSSENV 480
|||||
Qy 1152 RGSYNSSNNFQOPLTSRAEYCPWFEFETPAQPNAGRSVALPASSALSANKIAGPRKEETWD 1211
|||||
Db 481 RGSYNSSNNFQOPLTSRAEYCPWFEFETPAQPNAGRSVALPASSALSANKIAGPRKEEWD 540
1212 SFKV 1215
|||||
Db 541 SFKV 544
|||||

RESULT 2
Q9GKT3 ID Q9GKT3 PRELIMINARY; PRT; 317 AA.
AC Q9GKT3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 34.7 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREELLUM CORTEX;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
libraries.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AB052146; BAB19001.1; -
Hypothetical protein.
SEQUENCE 317 AA; 34746 MW; 26C0DB80505EFC4 CRC64;

Query Match 24.0%; Score 1532.5; DB 6; Length 317;
Best Local Similarity 93.4%; Pred. No. 2,7e-103;
Matches 297; Conservative 3; Mismatches 17; Indels 1; Gaps 1;

Qy 898 MLQKLSVIASAKETLGLAGKTQTAGVEERTKSKPLPKDKETNRHNSDNTETKDPQ 957
|||||
Db 1 MLQKLSVIASAKETLGLAGKTQTAGVEERTKSKPLPKDKETNRHNSDNTETKDPQ 60
|||||
Qy 958 PQNSNPAEPRKPQKSGIMKQORVNPPTTANSNLNPGTTQMKDNFDIGEYCPWEVYDLTPG 1017
|||||
Db 61 PQNSNPAEPRKPQKSGIMKQORVNPPTTANSNLNPGTTQMKDNFDIGEYCPWEVYDLTPG 120
|||||
Qy 1018 PVPSESKVKHYSIVASEMEKNPTSLSEKSHHKPKAAEVCOOSQKRIKAEVCLWESQ 1077
|||||
Db 121 PVPSESKVKHYSIVASEMEKNPTSLSEKSHHKPKAAEVCOOSQKRIKAEVCLWESQ 180
|||||
Qy 1078 GOSTLEDEKLLISKTPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOI 1137
|||||
```

```
Db 181 GOSTLEDEKHHFISKTPVLPERAKEENGOPRAANVCAGQSEELPPKAVASKTENENLNOI 240
Qy 1138 GHQEKKTSSSEENVGYSNNSNNFQOPLTSRAEYCPWFEFETPAQPNAGRSVALPASSALS 1197
|||||
Db 241 GHQEKKT-SPEENVGYSNNSNNFQOPLMSRAEYCPWFEFETPAQPNAGRSVALPVSSALS 299
|||||
Qy 1198 ANKIAGPRKEEWDSEFKV 1215
|||||
Db 300 ASKIAGPRKEEWDSEFKV 317
|||||

RESULT 3
Q9VR40 ID Q9VR40 PRELIMINARY; PRT; 669 AA.
AC Q9VR40;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG11923 PROTEIN.
GN CG11923.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chervy J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flooker A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
EMBL; AF003575; AAF50967.1; -
DR FlyBase; FBgn0031642; CG11923.
DR InterPro; IPR000337; -
DR Pfam; PF00003; 7tm_3; 1.
SQ SEQUENCE 669 AA; 75242 MW; FFE5721445DAB5F6 CRC64;
```

Query Match		9.0%; Score 571.5; DB 5; Length 669;
Best Local Similarity		26.1%; Pred. No.4.7e-33;
Matches 195; Conservative 113; Mismatches 316; Indels 123; Gaps		
QY	110 GKWPALASAHPSLHRALDTLTHTATNFLNVLMSNKSREONLQDDLDWQALVMSLLEGP	169
DB		
DB	6 GKWPC-RMYDYDTDAEDAROFIEFL-----SGKPPNANTPIAD-----EP	47
QY	170 SISRAAITTSOSLASAPQVFLQATREESRIILQDLSSAPHLA-----NATLETEWF	223
DB		
DB	48 --TRAEVSRRANGIAS-----YALNEDNLLAFAPASITHTVVKVFRDNVTIPDDQV	98
QY	224 HG---LRRKWRPHLRGRGNQGPRLGHSWRKRDGLGGDKSHFKSPVPLYECENSYKPG	280
DB		
DB	99 HNKAYLGYSVRE-----LGAAWNSTDGTQ-----EWGAPFRDCLNLTTRRWL	139
DB		
DB	281 WLVTLSAATYGLQPNLVPDFRGMVKVDINLQVDIDOCSSDGW---FSCTHKCHLNNSPC	337
DB		
DB	140 WPPRIISFS-----EHR-IKVVAAFIAADEVC--NDGLEEVFGRHCHCDRNTTFC	187
QY	338 MPTIKGL-GFVLGAYEICKAGYHPGVLPVNNFRRRQPDQHIISGTSKDYSE--EAYVCLUP	394
DB		
DB	188 LITENKPAATRDVYTCLCRESYYLPN-STLQGR-----GDRVELSEGVNYSCIP	237
QY	395 CREGCPCFADDSPC--FVQE-----DKYLRLAIISFOGLCMILLDFVMSLVVYHFRKAKSI	447
DB		
DB	238 CPGGCTNCDSNGVCLTFQEEVLNVDACURLJVAIVLGACILCCIVLGIVIVFRQKCKAI	297
QY	448 RASGLILLETILFGSLLLYFPVVVILYFEPSTERCILLRWALIGFATVYGTVTKLHRVL	507
DB		
DB	298 ASOMTVLETILLGIVLLYASVAVHFFPASTRECLLEPWLRELGFITCYGAILLKLYRHL	357
QY	508 KVELSTAQRIPYMTGGRVNRMLAVILLVYFWFLIGWTSVVCNLEK-QISLIGOGKTS	566
DB		
DB	358 VDFRTRKAHRW-VLRDQDVLKVLGTVMFVAVICYMAAFTASSLDLLESAQLESIRLEADT--	414
QY	567 HLIFNCLLDRWDYMTAVAEFLFLWGVLYCVAVRVPSPAFAEPYRMVAVAHNELIISAI	626
DB		
DB	415 ----NCHPLKMLVLTQTSMLTCLPGLHLSTASRNANTQFRERQFVLTALTEFLVSSS	470
QY	627 FHTIREVLASRLQSDWMLMLYFAHTLTVTVTIGLLLPK--FSHSSNNPRDDIATEAVE	684
DB		
DB	471 FYFLREVPYLPENSPSAILLALFIRSLTNSFALGLIFVPLKLVYQHKQGTSHD-----AG	524
DB	685 DELDMGRSSSYLNSINSAWSHSLDPPDIRDELKLYAQLEIYRKKKMITNNPHLQKKR	744
DB		
DB	525 QRUGGGYAGLCGLGDDPDIIGELTITSEMSPEDIRAEKRLYTOLEIMTKNKTLRQNP	584
QY	745 CSKKGLGRSIMRRIETPEITVSQRCSKEDKAGDHCTAG-----TALIRKNPDESSON	798
DB		
DB	585 GGRKAGHRFP-----SLQKSGKRLSAKHSNKHQHDIEITEAEPSTRPTEDSVSC	635
QY	799 TGKSKETEILK--NRVFSLKKSHTYDH	823
DB		
DB	636 SABGPTDTVAETISGVSHSMLSHSMVSH	662
RESULT 4		
ID	Q9VKA3	
QVKA3	PRELIMINARY;	PRT; 1677 AA.
OC	Q9VKA3;	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)	
DE	CG17215 PROTEIN.	
GN	CG17215.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephyridiodes; Drosophilidae; Drosophila.	
ON	NCBI_Taxid=7227;	
RX	[1]	

```

RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RC  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S. Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brannon K.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Beremand B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke S., Dahlen T.J., Hernandez J.R., Houck J.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA  Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA  Paizozolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reineert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA  "The genome sequence of Drosophila melanogaster.";
RT  Science 287:2185-2195(2000).
DR  PDBL: AE003635; AAF53175.1; -.
DR  Flybase: FBgn0032418; CG17215.
DR  InterPro: IPR000337; -.
DR  InterPro: IPR000834; -.
DR  InterPro: IPR002524; -.
DR  Pfam: PF01545; Cation_efflux; 1.
DR  PROSITE: PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
SQ  SEQUENCE 1677 AA; 188028 MW; F50A8D282A8E3B6E CRC64;

Query Match 4.8%; Score 307; DB 5; Length 1677;
Best Local Similarity 21.5%; Pred.No.2.8e-13;
Matches 134; Conservative 98; Mismatches 238; Indels 118; Gaps

QY 443 KAKSIRASGLILLETILFGSLLLYPPVILYFEPSTFRICLLRWLRLLGFAVTGVTLK 502
Db 1157 KAGRKGRSDWAI-----PPL-----DITWCIAKTRHMGFCITYSLLMK 1199
QY 503 -LHRVLKVFISRTAQRIPYTGVRWMRLAVILLVVFVFLIGWTSVQCNLEKQISLIGQG 562
Db 1200 TWRVSLTYRVKSAHKIK-LNDQOLLQWVPILLVLMVILGTWTISATPTAEVILD----- 1253
QY 563 KTSDFLLFNCLIDRWYMTAVAEFLFLLGVLGYLCYAVTPSAFHPRYMAVAHVNELI 622
Db 1254 --OSQLKFKQCSYNMWDHSLAIGEVEFFLAWGRVRCVYNRAESLYNEARLISAIYNTAL 1311
QY 623 IS---AIFTFIRFVLASRLQSDMLMLYFAHPLHVTVTIGLLIPKFSHSSNNPRDIIA 679
Db 1312 VNTAMVVFHVMFLPNAG---PDYKMLGEVRTLQTLSTTTTIALVFGPKILR-----V 1359
QY 680 TEAYEDELDMGRSGSYLNSINSANSEHSL-----DPEDIDELKKLYAQ 724

```

Query Match 4.8%; Score 307; DB 5; Length 1677;
Best Local Similarity 21.5%; Pred. No. 2.8e-13;
Matches .124; Conservative 98; Mismatches 238; Indels 118; Gaps 19;

[illegible]

Db 1360 FKQGDQKWDQ-----KAKVRSITASTSLNGVLVPESPDPOLYQENBELKEOVOKLAHQ 1412
QY 725 LEIYRKMKMTNNPHLQKRCCKGLGRSIRMRITEIPETVSRQCSKEDKAGDHGTAK 784
Db 1413 IEFMKTVMQMNRLK-----PKPGGYFITITSTFQAPYSKNTVSTAQTQTQKSDENSVKD 1468
QY 785 TALIRKNPESPNGTQSK--BETILKNRVFSLUKSHSYDHRVDQTERSSSLPTESQEE 842
Db 1469 CSI-----EIDGGGRTVLVDATE---FNF-----HLAD--TEKGLIVQEDERP 1509
QY 843 TTESNTLESLSCKKLTQKLKDESEAESESTVPLVCKKSASAHNLSEKKTGHPRTSMLQKS 902
Db 1510 SDEEAALIAQRRFLAPILDS-----LNLYYQLNDLDDTEHVRHIQTVAQ 1556
QY 903 LSVIASAKEKTGLAGKTQTACGVEERTKSKPLPKDKETNRNHSNDNTETKDPAPQNSN 962
Db 1557 MNDLTSSEETM-----VTVQNSPSPPPVGVVELLPISSDSST-----ASSSLY 1600
QY 963 PAEPRKQKSGI--MKQVRNPPTTANDSLNPGTQMK 998
Db 1601 AIHTPSAPHPSCVLLMQLNPLESPLLSGSDAITITQVE 1638
RESULT 5
ID Q9WIP3 PRELIMINARY: PRT: 264 AA.
AC Q9WIP3
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG18678 PROTEIN.
GN CG18678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov F.S., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarik C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003461; AAF47013.1; -.
DR FlyBase; FBgn0040664; CG18678.
DR InterPro; IPR000337; -.
DR Pfam; PF00003; 7tm.3; 1.
DR PRINTS; PR00248; GPCRMR.
SQ SEQUENCE 264 AA; 29532 MW; CAC5623C52703642 CRC64;
Query Match 4.4%; Score 281; DB 5; Length 264;
Best Local Similarity 28.5%; Pred. No. 1.5e-12;
Matches 65; Conservative 49; Mismatches 96; Indels 18; Gaps 5;
QY 472 LYFEPSTFCILLRWARLLGFATVYGVTVTLKLHRLVKVFLSSTAQRIPYMTGGRVYRMLA 531
Db 1 MYPNPLYTCTARIWLREIGFSLTYGALMLKLTWRTSVIFRVSRAKAVK-ITDAALLKRLG 59
QY 532 VILLVFWFLIGWTSVCONLEKQISL--IGCKTSDHLIFNMCLIDRWDMYTAVEFLF 589
Db 60 ICAGI-----GTCLLVRTLVSPPDVVGVVGTADDLKAFKLCTDMWDYFTTSNEVLF 110
QY 590 LIMGVYLCYAVRTVPSAFEPYMAVAVVHNLIIISAFTIRFVLASRLQSDWMLMLYFA 649
Db 111 LAWGVRLLCIMVKKAPSENESEFTSMAYINEFLTCTFNLVSMFLQSPANPDLLYIIFEC 170
QY 650 HTHLTVTVTIGLLL---IPKFSHSSNNPRDDIATAYEDELDMDGRSGS 694
Db 171 HTQLVTVELLALIFGSKIPTLTQVVIVLR---SGKSHQENIGMGTPKAS 215
RESULT 6
ID O45500 PRELIMINARY: PRT: 433 AA.
AC O45500;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE F39B2.8 PROTEIN.
GN F39B2.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z92834; CAB07388.1; -.
SQ SEQUENCE 433 AA; 49673 MW; B35F0A9C23DA096 CRC64;

Query Match 4.4%; Score 279.5; DB 5; Length 433;

RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shier B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA	Spie E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yen R.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT	"The genome sequence of Drosophila melanogaster";
RL	Science 287:2185-2195(2000).
RP	[2]
RN	SEQUENCE FROM N.A.
RA	Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA	Adayani A., Arcaina T.T., Baxter E., Blazej R.G., Butenhoff C.,
RA	Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA	Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA	Houston K.A., Humasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
RA	Park S., Sequeira A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,
RA	Celniker S.E.;
RT	"Full length Drosophila melanogaster cDNA sequence.";
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AE003736; AAF55916.1; -;
DR	EMBL; AF145639; AAD38614.1; -;
DR	FlyBase; FBgn0027575; BcDNA:GH07312.
DR	InterPro; IPR000337; -;
DR	InterPro; IPR001828; -;
DR	Pfam; PF000003; 7tm_3; 1.
DR	Pfam; PF01094; ANF_receptor; 2.
SO	SEQUENCE 1221 AA; 138123 MW; A57A9954F31FOA05 CRC64;
Query Match 4.0%; Score 254; DB 5; Length 1221;	
Best Local Similarity 17.9%; Pred.No.1.3e-09;	
Matches 213; Conservative 195; Mismatches 496; Indels 288; Gaps	
Qy	72 OKLAEEPMVDVASYLYTGDSHQLKRNACSG-RYELAGLPCKWPALASAHPSLHRALDTLT 130
Db	215 KKLREK--DVRITILGNFNHFARKAFCEAYKIDMYGRAYQLIMA-----T 259
Qy	131 HATFLNLVLMSQNSKRQNLODDLWDYQALWSLLGEGPISRAAITFTSDLSLAPAOV 190
Db	260 YSTDWNVTQDSCEVSE-----IATALEG-----ATLVDLLPLSTSGDIT 300
Qy	191 FLQATRESRIILLQDLSSAPHLANATLETFEPHGRLRK--WRPHLRRCPNGPGRLGH 248
Db	301 VAGITADEYLVEYDRUGT-----EYSRHFGTYTDIGMAALAIAQIVAB----- 344
Qy	249 SWRRKDLGGDKSHF-----KWSPPYLEC-ENGSIYPGWLVLTSSAIYGLQPNLVPEFRG 302
Db	345 --KREDLL---THFYDYRVKDWSVFLEALRNTSFEG---VTGPVRFYNNE----- 386
Qy	303 VMKVVDINLKQVDIDQCSDGWFSGTHTKLNNECPAPIKGLGVLGAYECICKAGFYHPG 362
Db	387 -RKANILINOLFQGMKEIGEYH-SQASHLDLSLGGPKVKWG----- 426
Qy	363 VLPVNFRRRQPQHISGSTKYDSEEAYVCLPCREGCPFCADSPCFVQEDK-----YLR 417
Db	427 -----KTPPK-----DTLIIEHSQNVPTIIV 450
Qy	418 LAISFQGLCMLLDFVMVLVVYHFRKAISRASGLILETLFGLSLLLYFPVVILFEPS 477
Db	451 SASASVGVIATVFLAFNIKY--RNQRYIKMSSPHLNNLIIVGCMMYLSIIIFGLDIT 508
Qy	478 -----TFFRCILLRWALLGFATVGTVTKLUHRVLKVLFSRTAQIPYMTGGRVMRL 530
Db	509 LSSVAAPFYICTARAWILTMAGFSLSFGAMFSKTRVHVSIFTDLKLNK-KVIDKYQLFMVV 567
Qy	531 AVILLVWFELIGH-TSSVCQNLKQISLIGCGKTSDHLLFNMCILDRWDYTAVAEFLF 589
Db	568 GVLLAIDAIAITTWQIADPYRETQKLEPLHHENIDDLVLPENECYSEHEMTFVSIY 627
Qy	590 -----LWGYLICAYVRTVP-SAFHBPRYMAYAVHNELLISAIFTIRFVLASRLOSOW 642

[illegible]

```
Db 628 AYKGLLVFAGLAWETRVISPALNDSKHIGFVSINVFITCLAGAAISLVLSDDRDLVF 687
Qy 643 MLMLYFAHRLHVTWTIGLLIPKFSHSSNNPRDDIATEAYEDELDMCRSGSYLNSINS 702
Db 688 VLLSFF--IIFCTATLCLVFPKVLKRNQGVVDRKVRATLPRMSKNGR----- 737
Qy 703 ANSEHSLDPEDIRDELKLYLAOLEYIKRKKMTN-NPHLQKRCCKKGLGRSIRMRITEI 761
Db 738 -----RDS-----SVCELEQRURDVKNTRCPRKALMEKENELQALIRKLG-- 778
Qy 762 PETVSRQ-----CSKEDKEDAGHTAKTALIRKNPPESSNTGKSEETLKNRVFSLK 815
Db 779 PE--ARKWIDGVTCTGSGNVGSELEPILNDDIVRLSAPPVRREMFSTTEVTMTSVDSVT 836
Qy 816 KSHSYDHY--VRDOTEESSLPTESQEBETEN-----STLESLSGKKLTKQLKEDSE 866
Db 837 STHVEMDNSFVSQSTVWAPSLPPKKKOSIVEHSHAPPTMMQPIQOOLQHLQHQHQ 896
Qy 867 AESTESVPLVCKSAHNLSEKKTGHPTSMLOKSLSVIASAKETLGLAGKTOTAG-- 924
Db 897 MQQHQ-----LQOQHQHQOQOQOQHHRLKRNVSVAQTDD---NIGSITSTACKR 948
Qy 925 -----VEERTK-----SKPLPKDKETNRHNSNSNTETKDPAPONSNAPEEP 967
Db 949 SGGDCSSMRERQSTASRHYDSGQSTPTARPKYSSHRNSNSTNISOSLSNMGPHSKP 1008
Qy 968 RKPQSGIMKQQRVNTPTANSILNCPCTOMKDNFDIGVCPWEVVDL-----TPGPVP 1020
Db 1009 STP-----AVIK-----TPTASDHRTSGNSALKNFVVSQSLDWTHTLSHAKORQSPRYA 1061
Qy 1021 SESKVQKHSIVASMEKNPT-----FSLKEKSHHKPAEAEVCOQSNOKRIDKAEVCLW 1074
Db 1062 SPQRCAEHGHGCHMTYDNTTSPIQRSVSEKRNKRHRPKP-----QKGTVC-- 1107
Qy 1075 ESQGSILEDEKLLSKTPVLPRAKEENGOPRAANCAGOSELPKAVASKTENENL 1134
Db 1108 QSETOSERDPPNPPQPCVQPRKVS-----RSSNIQHAHHSSPNVAPDK---QRS 1157
Qy 1135 NQIGHQEKK---TSSEENVRGYSNNSNFFQOPLTSRA-----EVCP 1173
Db 1158 RQRGQDSSIIYGASSETELLEGETAILPIFRKLLTEKSPNRYGRSAVGQSCP 1209

RESULT 8
Q9W1P4
ID Q9W1P4 PRELIMINARY; PRT; 176 AA.
AC Q9W1P4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DT CG18679 PROTEIN.
CG18679.
Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle E.G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottlie P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

```
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003461; AAF47012.1; -
DR FlyBase; FBgn0040663; CG18679.
DR InterPro; IPR000561; -.
DR SMART; SM00001; EGF-like; 1.
SQ SEQUENCE 176 AA; 19661 MW; 524E4E2C48CB7D33 CRC64;
```

Query Match 3.7%; Score 238.5; DB 5; Length 176;

Best Local Similarity 31.5%; Pred. No. le-09; Indels 33; Gaps 4;

Matches 52; Conservative 25; Mismatches 55; Indels 33; Gaps 4;

```
Qy 301 RGVKRVNINLQKVDIDQCSD-----GWFSCTHKCHLNNSCMPKIGLGVILGAYE 351
Db 9 RGTSGIDIDLRVDDIDQCQRHTPCTKRPLNIFACTDKCKQRTMCEAIMGLGFRGYSK 68
Qy 352 CTCRAGFYHGVLPVNNFRRGPDQIHSGSTKV-----SEBAYVCLPCRE 397
Db 69 CLCRKGFYFPDIVSQHKF-----FNGSLLEEYKMLGKNSYNSSEYECLPCAE 120
Qy 398 GCPFCADDSPCFVQEDKYLRLLAIIISFQGLCMLLDFVSLVYVHFR 442
Db 121 GDCSCDSSPCIAALNWPMTSILAL--ACIVIGLLPPAAWFTFR 163
```

RESULT 9

```
Q9VPS7
ID Q9VPS7 PRELIMINARY; PRT; 1305 AA.
AC Q9VPS7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG3022 PROTEIN.
GN CG3022.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle E.G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottlie P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agayani A., An H.-J., Andrews-Prannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chery J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jamal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003588; AAF51465.2; -
DR FlyBase: FBgn0031275; CG3022.
DR InterPro: IPR000005; -
DR InterPro: IPR000337; -
DR InterPro: IPR001828; -
DR Pfam: PF00003; 7tm_3; 1.
DR Pfam: PF01094; ANF_receptor; 2.
DR PRINTS: PR00248; GPCRGR.
DR PROSITE: PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
SQ SEQUENCE 1305 AA; 14358 MW; A9C385307650450 CRC64;

Query Match 3.6%; Score 227.5; DB 5; Length 1305;
Best Local Similarity 18.8%; Pred. No. 1.2e-07;
Matches 209; Conservative 171; Mismatches 381; Indels 351; Gaps 52;

Qy 181 DLSAP-APQVFLOATRESRILQDLSNAPH-----LANATLETFWHLRRKWRP 232
Db 383 ESMGAPWPDQRTACSNHLEQLAVENLIVSTHNSVGNVSYGLNHNMFNSQLRKOSA 442
Qy 233 HLHRRGNPGRLGHSW-----RRKDLGDKSHFKWSPPLYECNGSY 277
Db 443 QFH-----GODFGSGYGRISATATQSDRRRRRGVGTSGGHLFPEAISQYAPQTY 496
Qy 278 KPGMLVTLSSAIYGLQPNLVPFRGVKMDINLQKVIDOCSDGWFSGTHKCHLNSEC 337
Db 497 DAWAIALAL-----RAAEHW-----RRNEEQ 519
Qy 338 MPIKGLGFVLG--AYECIKAGFVH-PCVLPVNNFRRRGPDQIHSGST-----KDYSEE 388
Db 520 SKLGDFTYRSDMAWEFLQMGKHLFCVSGPVSF--SGPDR--VGTFATFQIQRGLEP 575
Qy 389 AYVCLPCREG----CPFC-----ADDSPCFVQEDKYLRLAIIISF 423
Db 576 VALYYPATDALDFRCPRCPVKHSGQVPIAKRVFKLRVATIAPL-----AFYTIATLSS 630
Qy 424 QGLCMLEDFVSMVYVHFRKAKSIRASGLJLITLFGSLLLYFPVVLVYFEPST-----478
Db 631 VGIALAIAFAFNL--HFRKLKALKLSPKLSNTAVGCIFFVATVILGLDHLSTLPSAE 688

Qy 479 --PRCILLRWARLL--GFATVYGTVTKLHRLVKVFLSRTAQ-----RIPYMTGG 524
Db 689 DSFATVCTARVYLLSAGFSLAFGSMFAKTYRVHRIF--TRTGSVFKDKMLQDILLLVGG 747
Qy 525 RVMRMLAVILLVWFVFLIGWTSVCONLEKQISLIGOGKTSDDLIF---NMCL---IDR 577
Db 748 ---LLIVDALLVLVWVTDPMERHLHNLTLIS-----ATDRSVVYQPOVEVCRSQHTOT 799
Qy 578 WDMYTAVEFLFLMGVYLCYAVRTP-SAFHEPRYMAVAVHNELIISAIFHTIRFVLAS 636
Db 800 WLSVLYAYKGLLVGVYMAWETRHVKIPALNDSQYIGSVSVITSAT-----VVULA 854
Qy 637 RLASDWMMLYFANHL---TVVTITGLLIPKFSHSSNNPRDDIATEAYEDELDMGRSG 693
Db 855 NLISERVTLAFITITLITLITSTTATLCLLPIK-----887
Qy 694 SYLNSINSAWSHS-LDP-----EDIRDELKLYAOLEIYKKKMI 734
Db 888 -----LHDIDWANDIIDPVIHSMGLAMECNTRRFVDDREIQYRVEQNRVTKKE---938
Qy 735 TNNPHLOKKRCKGLGRSIRMRITETPVSQCSKEDREGADH-----GTAKGT 785
Db 939 -----IQALDAEIRKLERLLESGLITTTTSSSTSL--LTGGHKLKPELTVTSGISQTP 991
Qy 786 ALIRKNPSSSGNTGKSEETLNRFSL-----KXSHSTYDHRVDQTESSSLPESQ 839
Db 992 AASKNRTPSISG-----ILPNLLSVLPVPIPRASWPSAEYMQIPMRRSVTFASQPQ 1043
Qy 840 EETTENSTLESLSGKKLTOKLKDSEASTESVPLVCKSAS-----AHNLSSEKKTG 892
Db 1044 LEE-----ACLPADQLINLRLAHQOATEAKTG 1070
Qy 893 -----HPRTSMLOKSLSVASAKEKTLGLAGKTOTAGVEERTKSQKPLPKDKETNRN 944
Db 1071 LINLRGIFSRSTSSNKG-STASLADQKGLKAAPKSH-MGLFTRL-----IPSSQATSCN 1123
Qy 945 HNSDNTETKDPQNPAPAEPRKPKOKSGIMKOORVNPPTANSDLNPCTOMKD-NF-D 1002
Db 1124 AIYNNPNQDSIPSEASHPNHLPKPIHRGSL-----TKSGTHLDHLT---KDPNPLP 1173
Qy 1003 IGEVCPMEVYDLPVPGVPVSESKVKQKHSIVASEM-----EKNPTF-----SLKEKSHHK 1051
Db 1174 IPTISGGEQDQTLG-----GKYVKLLETKVNFOLPSNRRPSVVQPPSRVRVRS 1225
Qy 1052 PK-----AAEVCQOSNOKRIDKAEVCLWESQGSILDEKLLISKTPVLPRAKEE--NG 1104
Db 1226 PRPHRLIPTCSLS-----ALAESDRP--GDSTSLSCSKSIPRISLQOATSG 1273
Qy 1105 GOPRA-----ANVCAGQSELPKAVASKTE 1130
Db 1274 GTWKSMTAGKSRSLGDSQEEQOAPANGTE 1305

RESULT: 10

076891
ID 076891 PRELIMINARY; PRY 5327 AA.
AC 076891;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE EG:49E4.1 PROTEIN.
GN FUTSCH OR EG:49E4.1 OR CG3064.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Prorygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Aphidoroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal x chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RESULT	15
Q17362	
ID	Q17362 PRELIMINARY; PRT; 6632 AA.
AC	Q17362;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE	UNC-89 PROTEIN.
DE	UNC-89.
GN	Cenorhabditis elegans.
OS	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2; MEDLINE=96180278; PubMed=8603916; Benian G.M., Tinley T.L., Tang X., Borodovsky M.; "The Cenorhabditis elegans gene unc-89, required for muscle M-line assembly, encodes a giant modular protein composed of Ig and signal transduction domains."; J. Cell Biol. 132:835-848(1996).
CC	- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX

Search completed: November 1, 2001, 10:19:21
Job time: 349 sec

|||||
481 ACAGCTGACCTGATATACGGGATGAATTGAAGAGCTGATGCA 432
724 GlnLeuGluIleTyrLysArgLysLysMetIleThrAsnAsnProHisLe 740
431 CAACGGAAGTCTACAAAGAAAGAGATGCTCGCCACACCCACACCT 382
740 uGlnLysLysArgCysSerLysLysGlyLeuGlyArgSerIleMetArgA 757
381 TCAAAAGAAACGCACTCAAGAAAGGCGCTGGGACGACCCCTGATGAAC 332
757 rGleThrGluIleProGluThr.....ValSerArgGlnCysSerLys 771
331 GCATCAGAGATCCAGAGACCATCCACCGGCGAGTGCAGCGCA 282
772 GluAspLysGluGlyAlaAspHisGlyThrAlaLysGlyThrAlaLeuI 788
281 GACGAC.....GGCAGCGAGCAGCGCAGCAACCGGGCACC.....TT 244
788 eArgLysAsnProProGluSerSerGlyAsnThrGlyLysSerLysGluG 805
243 GAAGAGAAACCACTTTGAGCCCAAGC...CACCATGGAAACCTCGGATG 197
805 luThrLeuLysAsnArgValPheSerLeuLysLysSerHisSerThrTyr 821
196 ACTCCCTTAAAAACAAAGGTCATGGTTTGCAGGAAGTCATCAGC...T 150
822 AspHisValArgAspGlnThrGluGluSerSerSerLeuProThrGluSe 838
149 GACCAGTTTGTGACCATGATCCAGAA..... 123
838 rGlnGluGluGluThrThrGluAsnSerThrLeuGluSerLeuSerGlyL 855
122ACTGGGAACCAAGACT...GAGTGTGTTCCCGGAG 92
855 ysLysLeuThr 858
|||||
91 ACAAAATGACT 81
seq_name: gb_gss3:CNS0320A

seq_documentation_block: 895 bp DNA GSS 15-MAY-2000
LOCUS CNS0320A
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
207B04 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL225307
VERSION AL225307.1 GI:7884197
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 895)
AUTHORS Roest-Crolius H., Jaillon O., Dasilva C., Fizes C., Fisher C.,
Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and
Weissenbach J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 895)
AUTHORS Roest-Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizes C., Wincker P., Brottier P., Quetier F.,
Saurin W. and Weissenbach J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 895)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES

source
1..895
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="207B04"
/clone_lib="G"
/notes="Genoscope sequence ID : COAG207DA02LP1-end : T7"

BASE COUNT 107 a 259 c 293 g 236 t
ORIGIN

alignment_scores:

Quality: 649.00 Length: 320
Ratio: 2.785 Caps: 14
Percent Similarity: 72.812 Percent Identity: 51.562

alignment_block:

US-09-775-181-2 x CNS0320A/rev ..

Align seg 1/1 to reverse of: CNS0320A from: 1 to: 895

731 LysLysMetIleThrAsnAsnProHisLeu.GlnLysLysArgCysSer. 746
|||||
895 AAGAAGATGCTGGCCACACACCCACCTCCCCAAGAAGCGGAGCTCCA 846
747 LysLysGlyLeuGlyArgSerIleMetArgIleThrGluIleProGlu 763
|||||
845 AAAAAGGTCCTGGCCGCTCGCTGATGAGCGCATCACCGATCCCCGA 796
763 tThrValSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisG 780
|||||
795 GTCGGTGACCCCAATCCAGCCGCGAGCAAGGAGGCTGCGACGACGC 746
780 lyThrAlaLysGly.ThrAlaLeuIleArgLysAsnProProGluSerSe 796
|||||
745 CAACCGCAACAGGTCTGCACGCTG...AAGAAGAACCCCTTTGATCCAC 699
796 rGlyAsnThrGlyLysSerLysGluGluThrLysLysAsnArgValPheS 813
|||||
698 CCACGCCCAACACCAACCAAGGAGGAGGTGTAAAGAGCAAGGTGTTCT 649
813 erLeuLysLysSerHisSerThrTyrAspHisValArgAspGlnThrGlu 829
|||||
648 CCCTGAAGAAGTCCACAGCAGCTACGATCAGTCCGTGACCAGGCCGAA 599
830 GluSerSerLeuProThrGluSerGlnGluGluGluThrThrGluAs 846
598 GGCCCGCATGGC.....TCAGCCCGCAGCAAGATGGAGGCTCCGGAAGG 555
846 nSerThrLeuGluSerLeuSerGlyLysLysLeuThrGlnLysLysG 863
554 CTCCCTCTCGGAGCGCTCATGGGCAAGAAGATGCCAAGATCCATCG 505
863 luAspSerGluAla.....GluSerThrGluSerValProLeuValCys 877
504 AAACATCAGTCTCCAGCGAGTCCACCGAGTCCGCTCTCTGCTGCTGC 455
878 LysSerAlaSerAlaHisAsnLeuSerSerGluLysLysThrGlyHisPr 894
454 AAGTCGGCAGCGCTCATAACTCTCGCGGAGCAAGAAGGCCATTCCTCCG 405
894 oArgThrSerMetLeuGlnLysSerLysSerValIleAlaSerAlaLysG 911
404 CAGAGCTCCATGCTGCAAGAGTCCCTCAGCGTCTATCGCTGGCGCAAG 355
911 luLysThrLeuGlyLeuAlaGlyLys.....ThrGlnThr 922
354 AGAGACCTCGGTTGACGCCAAAGGCCAGAACCTCAGAGGACACCCACA 305


```

AUTHORS  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
         keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
         Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
JOURNAL   scanning the human genome
MEDLINE   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT   99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3118 row: E column: 19
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 508.
FEATURES
source
1..508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-3118 Col=19 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT  166 a 129 c 109 g 98 t 6 others
ORIGIN
alignment_scores:
Quality: 597.00 Length: 140
Ratio: 4.489 Gaps: 0
Percent Similarity: 95.000 Percent Identity: 92.143
alignment_block:
US-09-775-181-2 x AQ764784 ..
Align seg 1/1 to: AQ764784 from: 1 to: 508
715 ArgAspGluLeuLysLysLeuTyrrAlaGlnLeuGluIleTyrrLysArgly 731
88 CAGGACGAGCTGAAATAAACTCTATGCCCACTGGAATATATAAAGAAA 137
731 slysMetIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysL 748
138 GAAGATGATCACAAACCCCACTCCAGAAAGCGGCTCGAAGA 187
748 ysGlyLeuGlyArgSerIleMetArgIleThrGluIleProGluThr 764
188 AGGGCTTANGTCGTCCATCATGACGACCATTCAGGAGATCCAGAGACA 237
765 valSerArgGlnCysSerLysGluAspLysGluGlyAlaAspHisGlyTh 781
238 GTCAGCGCGCAGTGTCTAAAGAGACAAAGGAGGCGCGCATGGCAC 287
781 rAlaLysGlyThrAlaLeuLeuArgLysAsnProProGluSerSerGlyA 798
288 ACCCAAGGCACTGCCCTCATCAGGAAGAACCTCCAGAGCTTCATGGA 337
798 snThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeu 814
338 ACACATGGAAATCCCAACGAGGAGACCCCTGAT.AACCGAGTCTCTCACTC 386
815 LysLysSerHisSerThrTyrrAspHisValArgAspGlnThrGluLys 831
387 AAGAATCCCAAGCACTATGACCACTGAGAGACCAACNACNAGAGTC 436
831 rSerSerLeuProThrGluSerGlnGluGluThrThrGluAsnSerT 848
|||||

```

```

437 CAGTAGCCTTACCACAGAAAGCCATGAGGAGAGACACAGA.AATTNCA 485

```

```

848 hrleuGluSerLeuSerGly 854
|||||

```

```

486 CACTGGAATCCNTGTCNGGT 505

```

```

seq_name: gb_gss15:AQ728799

```

```

seq_documentation_block:

```

```

LOCUS      AQ728799          543 bp      DNA          GSS          15-JUL-1999
DEFINITION HS.5464_B2.All.T7A.RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1040 Col=22 Row=B, DNA sequence.
ACCESSION  AQ728799
VERSION    AQ728799.1 GI:5500351
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 543)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE     Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1040 row: B column: 22
Seq primer: T7
Class: BAC ends
High quality sequence stop: 543.
FEATURES
source
1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1040 Col=22 Row=B"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT  168 a 139 c 124 g 107 t 5 others
ORIGIN

```

```

alignment_scores:
Quality: 511.00 Length: 141
Ratio: 4.223 Gaps: 0
Percent Similarity: 85.816 Percent Identity: 73.050

```

```

alignment_block:

```

```

US-09-775-181-2 x AQ728799 ..

```

```

Align seg 1/1 to: AQ728799 from: 1 to: 543

```

```

715 ArgAspGluLeuLysLysLeuTyrrAlaGlnLeuGluIleTyrrLysArgly 731
|||||

```

```

107 CAGCAGAGCTGAAAAAACTCTATGCCCACTGGATGTATTCAAGAA 156

```

```

731 sLysMetIleThrAsnAsnProHisLeuGlnLysLysArgCysSerLysL 748
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 GAGGATGATCTCAACACACCCCTCTCTGATTAAGCGTGCTCGAAGA 206
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
748 ysGlyLeuGlyArgSerIleMetArgArgIleThrGluIleProGluThr 764
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 AGGCGCTTAAGACGTTCCATCATGAGACTCATTACGGAGATCCCAAGACA 256
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
765 ValSerArgGlnCysSerLysGluAspGlyGluGlyAlaAspHisGlyTh 781
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 GACAGACGGCAATGCTCTAAAGAGGACAGAGAGGGCGCGACCATGGCAC 306
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
781 rAlaLysGlyThrAlaLeuLeuArgLysAsnProProGluSerSerGlyA 798
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
307 AGTCAAAATGCTCTCTCTCAATCAAGAGAAGAACCCCGGAGCTTTCACGGA 356
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
798 snThrGlyLysSerLysGluGluThrLeuLysAsnArgValPheSerLeu 814
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
357 ACACACCGAAATACAGAGGAGAGACCTGTAATAACCGAGTCTTCTACTC 406
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
815 LysLysSerHisSerThrTyAspHisValArgAspGlnThrGluGluSe 831
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
407 GAGAAATACACACAGCTCTTATGACCACTGAGAGACACACGGAAGACTC 456
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
831 rSerSerLeuProThrGluSerGlnGluGluGluThrThrGluAsnSerT 848
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 CTCGTACTACTCTCTCGAAAGCTCAGAGGAGGAGGAGCATCATTTGCA 506
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
848 hrLeuGluSerLeuSerGlyLys 855
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
507 CACTGTATCTCTCTGAGGAAA 529
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: gb_est63:BB354926

```

seq_documentation_block: 280 bp mRNA EST 12-JUL-2000
LOCUS BB354926 RIKEN full-length enriched, adult male corpus striatum Mus
DEFINITION BB354926 musculus cDNA clone C030003G02 3', mRNA sequence.
ACCESSION BB354926
VERSION BB354926.1 GI:9066754
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 280)
AUTHORS
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Horii,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
T., Tsunoda,Y., Watanabe,S., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)

```

```

TITLE
JOURNAL
COMMENT
Contact: Yoshhide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome.res@r.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

```

Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

```

source
1..280
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="C030003G02"
/location="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was combed with a primer [5' GAGAGAGAGAGATCCAGAGATCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATAATATCCCGCCCCCCCC 3']. cDNA was cloned into the xhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

```

BASE COUNT 79 a 65 c 62 g 74 t
 ORIGIN

```

alignment_scores:
Quality: 438.00 Length: 93
Ratio: 4.867 Gaps: 0
Percent similarity: 96.774 Percent Identity: 91.398

alignment_block:
US-09-775-181-2 x BB354926
Align seg 1/1 to: BB354926 from: 1 to: 280
633 ValLeuAlaSerArgLeuGlnSerAspTrpMetLeuMetLeuTyrPheAl 649
1 GTGCTAGCTTCGAGATTTCAGCCGCGATTCGATCTGATGCTGACTTGGC 50
649 aHisThrHisLeuThrValThrValThrIleGlyLeuLeuLeuLeuProL 666
51 ACATGCTTACTTGACTGTGACAGTTACCACTGGGCTGCTCTAATTCAA 100
566 ysPheSerHisSerSerAsnAsnProArgaspAspIleAlaThrGluAla 682
101 AGTTTTCACATTTCAAGCAATAATCCCGGAGAGACATTCGTACCAAGCA 150
683 TyrGluAspGluLeuAspMetGlyArgSerGlySerTyrLeuAsnSerSe 699
151 TATGAAGATGAGCTGGACATGGTTCGCTCTGGATCCTACCTGAACACGAG 200

```

699 r1leasnSerAlaTrpSerGluHisSerLeuAspProGluAspIleAArg 716
 201 TATCAATTGAGCTGGAGTGAACAGACGCTAGATCCAGAAGACATTTCGGG 250
 716 spGluLeuLysLysLeuTyAlaGlnLeu 725
 251 ATGAAGTGAATAAACTCTATGTCAGTTA 279

seq_name: gb_gss3:CNS02NOT

seq_documentation_block: 632 bp DNA GSS 14-MAY-2000
 LOCUS CNS02NOT Tetraodon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 151N10 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL205670
 VERSION AL205670.1 GI:7864489
 GSS: genome survey sequence.
 KEYWORDS Tetraodon nigroviridis.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 632)
 Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 632)
 Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 AUTHORS Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 TITLE Unpublished
 JOURNAL 3 (bases 1 to 632)
 REFERENCE Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES
 source

1..632
 Location/Qualifiers
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="151N10"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG151DG05LP1-end : T7"

BASE COUNT 99 a 206 c 219 g 99 t 9 others

alignment_scores:

Quality: 398.00 Length: 201
 Ratio: 2.707 Gaps: 7
 Percent Similarity: 73.134 Percent Identity: 46.766

alignment_block:

US-09-775-181-2 x CNS02NOT/rev ..

Align seg 1/1 to reverse of: CNS02NOT from: 1 to: 632

104 GluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSerAlaHisPr 120

630 GAGCTCCCTCTCTCGGGGAGKTCACCC.....GCGGCCCGCACGG 587

120 oSerLeuHisArgAla..LeuAspThrHisAlaThrAsnPheLeu 136

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

586 CTCGGTTACAGCGCGGCTGGAGCGCTCTGCACCGCCGCCAATTCTCTC 537
 137 AsnValMetLeuGlnSerAsnLysSerArgGluGlnAnsnLeu..GlnAspA 153
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 536 AGCGTGATCTCTCAAGCCAAAGGTCGCGAGAGCAGACTGCGGCGC...A 490
 153 spLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluGlyGluPro 169
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 489 ACATGGAGTGGTACCACGCTCTGGTTCGGAGCATCTGGAGGGGATTCC 440
 170 SerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuSerAlaPr 186
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 439 AGATATCCACGGGCTGTGTCTACATTTTCGGGCGCGATGCG...TCCGTGG 393
 186 oAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArgIleLeuL 203
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 392 GGRACCCGCGGTGCTCTCCAGCGACGAGCGCGCGCCAGCAGATCGTCC 343
 203 euGlnAspLysSerSerAlaProHisLeuAlaAsnAlaThrLeuGlu 219
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 342 TGCAGGACCTGTCCAGCGTGGCGCGCGGCTCMACAACCGCACCGCCAC 293
 220 ThrGluTrpPheHisGlyLeuArg...LysTrpArgProHisLeuHi 235
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 292 ACGGCTGTGTACACACGCCCAAGGAGCGCGAGAACGCCCGCTTCCG 243
 235 sArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrpArgA 252
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 242 GAAGAGGGTCTGAGCCAGAACCG...AGCCTAGATTCTGTGGCGCAGG 196
 252 rGlyAspGlyLeuGlyAspLysSerHisPhelystTrpSerProPro 268
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 195 GGGAGCCCGACGTGTTCGCGACAGRCGCGCAGTGAGGTGGTCCGCGCC 146
 269 TyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValThrLe 285
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 145 TACCTGGAGTGGCGGCGGCTTCHGCTCCACCGCTGGCTCTGACCTT 96
 285 uSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluPheArg 301
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 95 ATCRGCTGCCCTCTACGGCTCTGAAGGACGACCTGGCTCGGAGTTCA 47

seq_name: gb_gss3:CNS02LIV

seq_documentation_block:

LOCUS CNS02LIV 961 bp DNA GSS 12-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 227N18 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL178348
 VERSION AL178348.1 GI:7816405
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 961)

AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 961)

AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 961)


```

667 .....PheSerHi 669
303 CCGTTCGGGCTCTGAACCGTATCTGTGCCCTCCACCGATTCTCTCTT 352
669 sSerSerAsnAsnProArgAspAlaThrGluAlaThrGluAspG 686
.....
353 TGGCGGACCCACATCGGACGACATGGCTCGGAGGCTACGAGACG 402
686 luleuAspMetGlyArgSerGlySerTyrLeuAsnSerSerIleAsnSer 702
.....
403 AGCTGGACATGGGGCGCTCGGGTCTTACCTCAACGACGATCACGTCG 452
703 AlaTrpSerGluHisSerLeuAspProGluAspIleArg 715
.....
453 CGGTGGAGCGGACGACACCTGGATCCCGAAGACATTCGG 491

```

seq_name: gb_gss4:CNS03PB5

```

documentation_block:
LOCUS CNS03PB5 921 bp DNA GSS 17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
044B20 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL254426
VERSION AL254426.1 GI:7975438
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 921)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 921)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 921)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished

```

```

REFERENCE
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
AUTHORS Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
AUTHORS Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished

```

```

FEATURES
source
1..921
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="044B20"
/clone_lib="G"
/note="Genoscope sequence ID : C0B6044DA10SPI-end :
PUC-ori"
BASE COUNT 200 a 281 c 248 g 192 t
ORIGIN

```

```

alignment_scores:
Quality: 326.50 Length: 113
Ratio: 3.934 Gaps: 1
Percent Similarity: 73.451 Percent Identity: 57.522

```

```

alignment_block:
US-09-775-181-2 x CNS03PB5

```

Align seq 1/1 to: CNS03PB5 from: 1 to: 921

```

626 IlePheHisThrIleArgPheValLeuAlaSerArgLeuGlnSerAspTr 642
.....
478 CTGTTCTCTCTGACAGGTTCACTCTGTGTCAGGAGCTCCACCCAGACTG 527
612 pMetLeuMetLeuTyrPheAlaHisThrHisLeuThrValThrI 659
.....
528 GATGCTGCTGCTGTCTTTCACCCACACACACCTCACCGTCAACGTCACG 577
659 leGlyLeuLeuLeuLeuLeuProLys.....
578 TGGCCCTGCTGCTGCTGCCCAAGTGAGCGTCCGGTAGCTCTCTGTCCA 627
667 .....PheSerHi 669
628 CCGGTTCCGGGCTCTGAAACCGTATCTGTGCCCTCCACCGATTCTCTT 677
669 sSerSerAsnAsnProArgAspAlaThrGluAlaThrGluAspG 686
.....
678 TGGCGGACCCACATCGGACGACATGGCTCGGAGGCTACGAGACG 727
686 luleuAspMetGlyArgSerGlySerTyrLeuAsnSerSerIleAsnSer 702
.....
728 AGCTGGACATGGGGCGCTCGGGTCTTACCTCAACGACGATCACGTCG 777
703 AlaTrpSerGluHisSerLeuAspProGluAspIleArg 715
.....
778 CGGTGGAGCGGACGACACCTGGATCCCGAAGACATTCGG 816

```

seq_name: gb_est64:BB429778

```

seq_documentation_block:
LOCUS BB429778 288 bp mRNA EST 18-JUL-2000
DEFINITION BB429778 RIKEN full-length enriched, adult male hippocampus Mus
musculus cDNA clone C630030A14 3', mRNA sequence.
ACCESSION BB429778
VERSION BB429778.1 GI:9269505
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 288)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukui,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Kai,C., Kawai,J., Kikuchi,N.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length

```


GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:14:19 ; Search time 45:13 Seconds
(without alignments)
804.648 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MGAWAYPELLCLLQAQLGLG.....YMTAVGMWSLVSDGLTIFQ 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	141.5	4.4	974	21	AA45000 Human calcium sens
2	139.5	4.4	917	20	AA45000 Chimeric hCAR/hmgl
3	139.5	4.4	1276	20	AA45000 phCAR/hmglur2*Gai5
4	136	4.3	901	21	AA45001 Human calcium sens
5	133.5	4.2	1026	18	AAW32059 Dogfish shark kidn
6	133.5	4.2	1079	19	AAW54847 Rat kidney calcium
7	133.5	4.2	1079	19	AAW38275 Rat kidney cell ca
8	133.5	4.2	1079	20	AA41781 Rat parathyroid ca
9	133.5	4.2	1079	20	AAW94928 Rat kidney extrace
10	133.5	4.2	1079	20	AAW89566 Rat parathyroid ca
11	133.5	4.2	1079	21	AA451828 Rat calcium recept

12	132	4.1	1001	21	AA44999 Human calcium sens
13	127.5	4.0	840	21	AA45021 Rat sensory transd
14	127.5	4.0	840	21	AA45026 Rat sensory transd
15	127.5	4.0	840	21	AA45027 Rat sensory transd
16	127.5	4.0	840	21	AA45028 Rat sensory transd
17	127	4.0	1078	17	AAW11889 Parathyroid calciu
18	127	4.0	1078	19	AAW54846 Human parathyroid
19	127	4.0	1078	19	AAW38274 Human parathyroid
20	127	4.0	1078	20	AAW28840 Human calcium rece
21	127	4.0	1078	20	AA41780 Human parathyroid
22	127	4.0	1078	20	AAW89565 Human parathyroid
23	127	4.0	1078	21	AA451827 Human calcium rece
24	127	4.0	1078	21	AA470325 Human wild type ca
25	123	3.8	1085	19	AAW54844 Bovine parathyroid
26	123	3.8	1085	19	AAW38272 Bovine parathyroid
27	123	3.8	1085	20	AA41778 Bovine parathyroid
28	123	3.8	1085	20	AAW89563 Bovine parathyroid
29	123	3.8	1085	21	AA451825 Bovine calcium rec
30	120	3.8	824	20	AA49124 Human mglur2 prote
31	120	3.8	872	17	AAW89580 Metabotropic gluta
32	120	3.8	1035	20	AA49128 Chimeric hmglur2/h
33	120	3.8	1394	20	AA49129 pmglur2/CaR*Galpha
34	120	3.8	1397	20	AA49134 pmglur2/CaR*Galpha
35	119.5	3.7	1088	17	AAW11888 Parathyroid calciu
36	119.5	3.7	1088	19	AAW54845 Human parathyroid
37	119.5	3.7	1088	19	AAW38273 Human parathyroid
38	119.5	3.7	1088	20	AA41779 Human parathyroid
39	119.5	3.7	1088	20	AAW89564 Human parathyroid
40	119.5	3.7	1088	21	AA451826 Human calcium rece
41	118.5	3.7	1058	18	AAW25762 Amino acid sequenc
42	118	3.7	872	17	AAW95052 Human metabotropic
43	116	3.6	906	16	AAW64250 Human mglur1B. Ho
44	116	3.6	1056	18	AAW25765 Amino acid sequenc
45	116	3.6	1194	14	AAW42199 HSmglur1. Homo sa

ALIGNMENTS

RESULT 1
AA45000
ID AA45000 standard; Protein; 974 AA.
XX
XX AA45000;
XX
XX 20-JUN-2000 (first entry)
XX
XX Human calcium sensing receptor isoform, CaSRd.

XX Calcium Sensing Receptor; CaSRd; isoform; human; splice variant;
KW hypotensive; osteopathic; receptor activity; calcium level; modulator;
KW treatment; hyperparathyroidism; osteoporosis; Paget's disease;
KW hypercalcaemia malignancy; hypertension; gene therapy.

XX Homo sapiens.
OS Synthetic.
XX
XX WO200006601-A1.

XX
XX 10-FEB-2000.

XX
XX 28-JUL-1999; 99WO-US17116.
XX
XX 30-JUL-1998; 98US-0094702.

XX
XX (AVET) AVENTIS PHARM PROD INC.

XX
XX Yu KT, Labaudiniere RF, Thrower LW;
XX WPI: 2000-195263/17.

XX
XX N-PSDB; AA50616.

XX
XX Nucleic acids encoding isoforms of human calcium sensing receptor for

PT treating, e.g. hyperparathyroidism or osteoporosis

PS Claim 24; Page 66-69; 81pp; English.

CC The present amino acid sequence is an isoform of the human calcium
CC sensing receptor CasR, expressed in the kidney. It is a splice variant
CC of the wild type CasR, that arise from deletion of amino acids
CC 361-465, from the extracellular domain, that comprise acidic residues.
CC It has hypotensive and osteoprotic activity. The CasR isoforms can be
CC used to identify agonists and antagonists that modulate the receptor
CC activity and calcium levels. These modulators are useful for treating
CC hyperparathyroidism or osteoporosis. Paget's disease, hypercalcaemia
CC malignancy or hypertension. The DNA sequence is also useful for altering
CC the CasR activity and in gene therapy.
CC Note: This sequence has deletion of amino acids 361-465, according to
CC the sequence shown in the specification, but has been stated as
CC 358-462 in the claims.

XX Sequence 974 AA;

Query Match 4.4%; Score 141.5; DB 21; Length 974;
Best Local Similarity 19.0%; Pred. No. 0.00023;
Matches 114; Conservative 89; Mismatches 246; Indels 151; Gaps 26;

QY 75 AEEVPMVASLYLTGDSHQLKRANCSGRYELAGLPKWPALASHPULRALDTLTHATN 134
D 132 sehlpstiaavvgatgsvstavanllgfyf-----pqvsyasss--rllsnknqfks 182
QY 135 FLNWMLOSNSREQNLQDLDMYQALVWSLL-----EGEPSTSRRAITFSDLSAP 186
D 193 flrt-ipndehqatamadllefyf--wnvgtiaadddygrgplekfreaeardicid 238
QY 187 APQVFLQATRESRIQLQDLSSAPHLANATLETFEWFHGLRRKWRPHLHRRGPNQGP--- 243
D 239 fseilqsydsde-----elqhvevignstakv-----ivfssgpdleplik 281
QY 244 ---RGL-GHSWRKDLGGDKSHFKNS-----PPYLECENSGYKPGWLVLTSSAIYG 291
D 282 elvrnritgkiwasea-----wasssliampqyfhvvggtig-----fa 321
QY 292 LQPNLVPEFRGVKVDINLQKVDIDQSSDGFSGTHKCHL-----NN-----SECM 338
D 322 lkaqigpgrflek-kvhpkrsvingfakew-eetfnchlrhlnftnmgeqvtfdcc 379
QY 339 PIKGLGFVL-----GAYECICKAG---FYHPGVLPVNNRRRRGPDQH-- 377
D 380 dlvgnyslinwhlspedgsivfkevgynvyakgerlfineekilwsgfrevpfns 439
378 ---ISGSTKOVSE---AYVCLPCREG-----CPCADD-----SPCFVQED 413
D 440 rdcilagrtrkgliegeptccfecvcpdgyesdetdasacnkcpcdfwsnhtsciakei 499
QY 414 KYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIRASGLILLETILFGSLLLYF 467
D 500 eflswtepfgiatlifavglfcafvlgvfkfntpivkatnrelyillfllccfs 559
QY 468 PVVILYFEPSTFRCILLRWARLLGFATVYGTVTLKLRHLVKVFLSRTAQRIPYMTGGRVM 527
D 560 ssiffigeqdwtrclqpafgsvfiscilvktvrvllvfeakipstfhrkwglnl 619
QY 528 RMLAV-----ILLVFWFLIGHTSSV-CONLEKQISLI-----GQKQTSDDLIFNCLLI 575
D 620 qllivlctfmqlvicvliwlytappssyrnrgeladeifitcheqslmaigfligytcll 679

RESULT 2

AA49126

ID AA49126 standard; Protein; 917 AA.

XX

AC AA49126;

XX

DT 07-JAN-2000 (first entry)

XX Chimeric hCAR/hmGluR2 protein sequence.

DE XX G-protein fusion receptor; Car; calcium receptor; GluR; head injury;
KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
KW cognitive disorder.

OS Homo sapiens.

XX WO951641-A1.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSP-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;
PI Simin RT;

XX WPI; 1999-610995/52.

XX N-PSDB; AAZ31057.

XX New G-protein fusion receptors and chimeras containing domains from
PT different receptors, used to screen for modulators, potentially useful
PT e.g. for treating or preventing stroke or Alzheimer's disease -
XX Disclosure; Fig 10; 255pp; English.

XX The invention relates to G-protein fusion receptors (I) comprising:
CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
CC intracellular (ICD) domains, each chosen independently from a Car
CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
CC the C-terminus of ICD; and (3) a G-protein (Gp) linked to ICD or the
CC linker. (I), and recombinant chimeric receptors (CR) without the Gp
CC component, are used to assess function of the various domains and to
CC identify compounds (e.g. allosteric modulators or antagonists) that act
CC on these domains. The modulators are potentially useful for treating or
CC preventing diseases associated with the receptors, e.g. stroke, head or
CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
CC (1) for recombinant production of corresponding proteins; and (2) to
CC produce cells used in screening for modulators. Use of Car and mGluR
CC domains allows presentation of GABABR domains, to a binding agent, in a
CC form more like the natural domain structure compared with use of
CC incomplete receptors, lacking one or more domains. By shuffling different
CC domains, agents can be identified that affect particular domains of a
CC receptor.

XX Sequence 917 AA;

Query Match 4.4%; Score 139.5; DB 20; Length 917;
Best Local Similarity 24.2%; Pred. No. 0.00032;
Matches 68; Conservative 43; Mismatches 111; Indels 59; Gaps 15;

QY 311 QKVVIDQSSDGFSGTHK---HLNNS-CMPKIGLGVILGAYECICKAG---FYHPGV 363
D 475 eqvtfdcc---gdlvgnyslinwhlspedgsivfkevg---ynvyakgerlfineek 527
QY 364 LPVNNRRRRGPDQH-----ISGSTKOVSE---AYVCLPCREG-----CPCAD 404
D 528 ilwsgfrevpfnsrclagrtrkgliegeptccfecvcpdgyesdetdasacnkcpc 587
QY 405 D-----SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIR 448
D 405 D-----SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIR 448

QY 397 EG-----CPFCADD-----SPCFVQEDKYLK-----LAIISFQGLCMLLDFV 433
 Db 577 egefsdendasactkcpndfwsnhtsciakeyleylswtepfialtifavilgilitf 636
 QY 434 SMLVVVHFRKAKSTRASGLILLETILFSGSLLYFPVVLVFEPTFCILLRWARLLGFA 493
 Db 637 vlgvfikrntpvkatnrelsylllflsicfssllfigepdrwtcrlrqafglsfv 696
 QY 494 TVYGTVTLLKLRVFLSRTAQRIPTYMTGGRVMRLAVILLY-----VFVFLIGWTS 546
 Db 697 lciscilvknrvllvfeakiptslhrkwwglnqlfllvflcivlqvltciliwlytapps 756
 QY 547 SVCQNLKQISLIGQKTSDFHFLFNCLDRWDMYTAGWMSLVSY 592
 Db 757 sy-rnhele-----devifitc-----degslmaigfligy 786

RESULT 6

AAW54847
 AAW54847 standard; Protein; 1079 AA.

AAW54847;

01-SEP-1998 (first entry)

Rat kidney calcium receptor 3A protein 4Kb fragment.

Calcium ion concentration; parathyroid hormone; homeostasis; kidney;
 calcium receptor; detection.

Rattus sp.

US5763569-A.

09-JUN-1998.

07-JUN-1995; 95US-0484565.

07-JUN-1995; 95US-0484565.

23-AUG-1991; 91US-0749451.

11-FEB-1992; 92US-0834044.

21-AUG-1992; 92US-0934161.

12-FEB-1993; 93US-0017127.

23-FEB-1993; 93US-0009389.

22-OCT-1993; 93US-0141248.

19-AUG-1994; 94US-0292827.

21-OCT-1994; 94WO-US12117.

08-DEC-1994; 94US-0353784.

(BGHM) BRIGHAM & WOMENS HOSPITAL.
 (NPSP-) NPS PHARM INC.

Brown EM, Garrett JE, Hebert SC;

WPI; 1998-347412/30.

N-PSDB; AAW26965.

Calcium receptor poly:peptide(s) - useful for drug screening or
 antibody production

Claim 8; Fig 50; 174pp; English.

The tissue from which the rat kidney calcium receptor and receptors from
 bovine parathyroid and rat kidney are derived, respond to changes, and
 control changes, in calcium ion concentration, e.g. parathyroid hormone
 regulates Ca²⁺ homeostasis in blood and extracellular fluid, and kidney
 function alters through changes in Ca²⁺ levels in juxtaglomerular and
 proximal tubule cells in the kidney. The purified receptors (produced
 recombinantly) can be used to screen for compounds that modulate calcium
 receptor activity, especially those that can be used to treat diseases
 associated with the receptors in these tissues. They can also be used
 to raise antibodies for use in detection assays.

XX

XX SQ Sequence 1079 AA;

Query Match 4.2%; Score 133.5; DB 19; Length 1079;

Best Local Similarity 21.7%; Pred. No. 0.0016;

Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGPNQGRGLGHSRRKDCGLGDKSHFKWSPPYLECENGSYKPGWLVTLSAIVGLQP 294

Db 377 heeg---gnrlsnstafplctgdeninsvetpymdyehlrisyvylavysiahald 433

QY 295 --NLVPEFRGVMKVDINLQKVIDQCSSDGFSGTHKCHLNNS-----ECMPIKG 342

Db 434 iytclpg-rglf---tngscadikkv--eawqlklrhlrlnftnnmgeqvtfdccgdlvg 487

QY 343 LGFVL-----GAYECICKAG---FVHPGVLPVNNFRRRGPDOHIS--- 379

Db 488 nysiinwhlspedgsvfkevgyynvyakgerlfinnekilwsgfsrevpfncsrdcq 547

QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYLK 417

Db 548 agtrkgliageptccfecvcpdgyesgetdasackpddfwsnhtsciakeiefla 607

QY 418 -----LAIISFQGLCMLLDFVSMVLYVHPRKAKSTRASGLILLETILFSGSLLYFPVVI 471

Db 608 wtepfialtlfavlglftafvlgvfikrntpvkatnrelsylllflsicfssllf 667

QY 472 LYFEPSTFCILLRWARLLGFAIVYGTVTLLKLRVFLSRTAQRIPTYMTGGRVMRLA 531

Db 668 figepqdwtrlrqafglsfviciscilvktorvllvfeakiptsfhrkwwglniqfll 727

QY 532 V-----ILLVVFVFLIGWTSSVCQNLKQISLIGQKTSDFHFLFNCLDRWDMYTAG 584

Db 728 vflctfmqlliciwiptyappssy-rnhele-----deiifitc---hegsimal 773

QY 585 GMSLVSY 592

Db 774 g---sligy 779

RESULT 7

AAW38275

ID AAW38275 standard; Protein; 1079 AA.

XX AC AAW38275;

XX DT 08-MAY-1998 (first entry)

XX Rat kidney cell calcium receptor 3A (RakCar 3A).

XX DE Rat kidney cell calcium receptor 3A; RakCar 3A;

XX KW calcium homeostasis; hyperparathyroidism; osteoporosis.

XX OS Rattus sp.

XX US5688938-A.

XX PN 18-NOV-1997.

XX PD 07-JUN-1995; 95US-0485588.

XX PR 07-JUN-1995; 95US-0485588.

XX PR 23-AUG-1991; 91US-0749451.

XX PR 11-FEB-1992; 92US-0834044.

XX PR 21-AUG-1992; 92US-0934161.

XX PR 12-FEB-1993; 93US-0017127.

XX PR 23-FEB-1993; 93US-0009389.

XX PR 22-OCT-1993; 93US-0141248.

XX PR 19-AUG-1994; 94US-0292827.

XX PR 21-OCT-1994; 94WO-US12117.

XX PR 08-DEC-1994; 94US-0353784.

XX Screening for calcium receptor-active compounds - by recombinant
PT expression of nucleic acid encoding calcium receptor and determining
PT the effect of compounds on calcium receptor activity
XX
XX Claim 1; Fig 50; 176pp; English.
XX
XX A method has been developed of screening for a compound able to affect
CC one or more activities of a calcium receptor (CR) comprises: (A)
CC contacting a recombinant cell with a test compound, where the
CC recombinant cell comprises a recombinant nucleic acid expressing the CR,
CC provided that the cell does not have functional CR expression from
CC endogenous nucleic acid; (B) determining the ability of the test
CC compound to affect one or more activities of the calcium receptor; and
CC (C) comparing the ability with the ability of the test compound to
CC affect the one or more CR activities in a cell not comprising the
CC recombinant nucleic acid. The present sequence represents rat
CC parathyroid CR, designated a prakCar 3A. The nucleic acid sequence of
CC prakCar 3A can be used as part of the recombinant nucleic acid in the
CC method described above. The compounds identified can be used to treat
CC diseases or disorders characterised by abnormal calcium homeostasis, e.g.
CC hyperparathyroidism, osteoporosis and other bone and mineral-related
CC disorders. They can also be used for the treatment of diseases and
CC disorders associated with disrupted Ca²⁺ responses, e.g. seizures,
CC stroke, spinal cord injury, hypoxia-induced nerve cell damage such as in
CC cardiac arrest or neonatal distress, epilepsy, neurodegenerative
CC diseases such as Alzheimer's disease, Huntington's disease and
CC Parkinson's disease, dementia, muscle tension, depression, and anxiety.
XX
XX Sequence 1079 AA;
SQ

Query Match 4.2%; Score 133.5; DB 20; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.0016;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;
Oy 235 HRRGNQPGRLGHSWRKDGKSHFKWSPDYCEGNSYKPGWLVLTSSAIYGLQP 294
Db 377 heeg---gnrlinsstafrpctgdeninsvetyehrlisynvylavysahahql 433
Oy 295 --NLVPERGVKVDINLOKVDIQCSSDGFSGTHKCHLNNs-----ECMPKIG 342
Db 434 iytclpg-rgif---tngscaidkkv--eawqvklhrlhntnmgeqvtfcdcgdlvg 487
Oy 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNRRRRQDQHS----- 379
Db 488 nyslinwhlspedgsivfkegyynvyakgerlfineekilwsgfrevpfscsrdoq 547
Oy 380 -GSTKDVSEE-----AYVCLPREG-----CPFCADD-----SPCFVQEDKYL 417
Db 548 agtrkgilegeptccfecvecpdgsgetdasackcpddfwennhtsciakelefla 607
Oy 418 -----LAIISFOGLCMLDVPVLMVYVHFRKASIRASGLILLETILFGSLLYFPVWI 471
Db 608 wtefgtaiftfavgifltafvgfkfntpvkatnrelsyllflslccfssif 667
Oy 472 LYFPSPFRCLLRWARLLGPATVYGVTVTLKHLRVLVFLSRQAQIPYMTGGRVMRLA 531
Db 668 figeqqdwtrlcpafigsvlscilvktntnrvllvfeakiptsfhrkwwgnlqlfll 727
Oy 532 V-----ILLVWFELIGWTSVCQNLEKQISLIGOKTSYDHLIFNMCLIDRWYMTAV 584
Db 728 vflctfmqilicliwlycapssy-rnhele-----delifitc-----hegsimal 773
Oy 585 GMSVLSVY 592
Db 774 g--sligy 779

RESULT 11
AAY51828
ID AAY51828 standard; Protein; 1079 AA.
XX

AC AAY51828;
XX 09-JUN-2000 (first entry)
XX Rat calcium receptor prakCar3A protein.
XX
XX Calcium receptor; treatment; calcimimetic; calcilytic; osteopathic;
KW cerebroprotective; cytostatic; neuroprotective; dermatological;
KW tranquilizer; vulnerary; antiulcer; immunosuppressive; hypotensive;
KW cardiant; parathyroid hormone; osteoporosis; calcitonin secretion;
KW hyperparathyroidism; Paget's disease; rat.
XX Rattus sp.
XX OS
XX US6031003-A.
XX 29-FEB-2000.
XX 07-JUN-1995; 95US-0484719.
XX 23-AUG-1991; 91US-0749451.
PR 11-FEB-1992; 92US-0834044.
PR 21-AUG-1992; 92US-0934161.
PR 12-FEB-1993; 93US-0017127.
PR 22-FEB-1993; 93US-0009389.
PR 22-OCT-1993; 93US-0141248.
PR 19-AUG-1994; 94US-0292827.
PR 21-OCT-1994; 94WO-US12117.
PR 08-DEC-1994; 94US-0353784.
XX (NPSP-) NPS PHARM INC.
PA (BCHM) BRIGHAM & WOMENS HOSPITAL.
XX Balandrin MF, DelMar EG, Moe ST, Nemeth EF, Van Wagenen BC;
PI N-PSDB; AA289299.
XX WPI: 2000-301969/26.
DR N-PSDB; AA289299.
XX
XX Treating disorders or diseases in a patient by modulating inorganic ion
PT receptor activities especially calcium receptor for
PT hyperparathyroidism, by administering a calcimimetic or calcilytic
PT compound -
XX
XX Example 30; Column 155-162; 194pp; English.
XX
XX This invention describes a novel method of treating disorders by
CC modulating calcium receptor activity in vitro by administering a
CC calcimimetic, or a calcilytic compound. The products of the invention
CC have osteopathic, cerebroprotective, cytostatic, neuroprotective,
CC dermatological, tranquilizer, vulnerary, antiulcer, immunosuppressive,
CC hypotensive and cardiant activity. The method is suitable for reducing
CC parathyroid hormone level in a patient to that of a normal individual,
CC treating a patient having osteoporosis, to inhibit bone resorption, and
CC to stimulate calcitonin secretion in vitro or in vivo. The level of
CC parathyroid hormone is reduced to cause a decrease in plasma Ca²⁺. The
CC method is useful in treating disorders in humans such as
CC hyperparathyroidism, Paget's disease and osteoporosis. Also for treatment
CC or prevention, based on the affected cells, of other disorders and
CC conditions like seizures, stroke, head trauma, spinal cord injury,
CC hypoxia-induced nerve cell damage such as in cardiac arrest or neonatal
CC distress, epilepsy, Alzheimer's disease, Huntington's disease,
CC Parkinson's disease, dementia, muscle tension, depression, anxiety,
CC panic disorder, OCD (not defined), post-traumatic stress disorder,
CC schizophrenia, neuroleptic malignant syndrome and Tourette's syndrome,
CC diseases involving excess water reabsorption by the kidney such as
CC syndrome of inappropriate ADH secretion (SIADH), cirrhosis, congestive
CC heart failure and nephrosis, hypertension, preventing and/or decreasing
CC renal toxicity from cationic antibiotics (e.g. aminoglycoside
CC antibiotics), gut motility disorders such as diarrhoea and spastic
CC colon, GI (gastrointestinal) ulcer diseases, GI diseases with excessive
CC calcium absorption such as sarcoidosis and autoimmune diseases and
CC organ transplant rejection. This sequence represents the rat calcium
CC receptor prakCar3A which is described in the method of the invention.

RAY45021
ID AAY45021 standard; Protein; 840 AA.
XX AC AAY45021;
XX DT 31-MAY-2000 (first entry)
XX DE Rat sensory transduction G-protein coupled receptor-B3.
XX KW Rat; sensory transduction; G-protein coupled receptor-B3; GPCR-B3;
XX KW sensory cell; taste receptor cell; screen; taste modulator;
XX KW pharmaceutical; food industry; taste topographic map; tongue.
XX OS Rattus sp.
XX FH Key Location/Qualifiers
XX FT Domain 1..580
XX FT /label= Extracellular_domain
XX PN WO200006592-A1.
XX DE 10-FEB-2000.
XX PF 27-JUL-1999; 99WO-US17099.
XX KW 28-JUL-1998; 98US-0094465.
XX PR (REGC) UNIV CALIFORNIA.
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX PI Zuker CS, Adler JE, Lindemeier J, Ryba N, Hoon M;
XX WPI; 2000-205451/18.
XX DR N-PSDB; AAZ50743.
XX DE New isolated sensory transduction G-protein coupled receptor, useful
XX PT for developing products for use in studying and modulating the taste
XX PT transduction pathway -
XX PS Claim 22; Page 75; 83pp; English.
XX CC The present sequence is a taste cell specific G-protein
XX CC coupled receptor, GPCR-B3 which is involved in sensory transduction.
XX CC This sequence was isolated from the IZAP rat circumvallate cDNA library.
XX CC GPCR-B3 is specifically expressed in foliate and fungiform cells, with
XX CC lower expression in circumvallate taste receptor cells of the tongue.
XX CC The protein has an extracellular domain, seven transmembrane
XX CC domains and an intracellular domain.
XX CC The present sequence is used to screen compounds that modulate sensory
XX CC signalling in taste cells, especially taste modulators useful in
XX CC pharmaceutical and food industries to customise taste. The sequence
XX CC can also be used as probe for identifying taste cells and
XX CC subsets of taste receptor cells such as foliate, fungiform and
XX CC circumvallate. Such probes are also useful to generate taste
XX CC topographic maps that elucidate the relationship between the taste
XX CC cells of the tongue and sensory neurons leading to taste centres
XX CC in the brain.
XX SQ Sequence 840 AA;

Query Match 4.08; Score 127.5; DB 21; Length 840;
Best Local Similarity 22.24; Pred. No. 0.0041;
Matches 60; Conservative 48; Mismatches 89; Indels 73; Gaps 14;

OY 306 VDINLQK-----VDIDOCSSDGFSCFTHKCHLNSECMPIKGLGVLAGYECI-CK 355
DB 475 idnktkiqwhgknqpvsvctcd-claghrhvvvghhc-----cfcvpcpe 522
OY 356 AGFYHGPVLPVNNRRRGPQDHQISGTSKDVSEAYVCLPC--REGCPFCADSDPCFVQED 413
DB 523 agtf-----ln-----mselhicqpcgcewap--kesttcfrtv 556

OY 414 KYL-----RLAIIISFOGLCMLDGF-VSMLVVYHFRKAKSTRASG-----LILLETILFGS 462
DB 557 eflawhepislviaaantllllllivtagllfawhfh-tpvvrsggricfclmglsvags 615
OY 463 LLLYFPVILYFEPSTFRCILLRWARLIGFATVYGTVLKLRHLVKVFLSTAQRIPLYMT 522
DB 516 csfy-----sffgeptvpacilrqlpfsigfaifscitrsqfvlifkftskvptfyr 671
OY 523 -----GGRVMRLAVILLVVFVFLIGWT 545
DB 672 waqnhgaglfvsvstvhllcltvlwmwt 701

RESULT 14
AAY45026
ID AAY45026 standard; Protein; 840 AA.
XX AC AAY45026;
XX DT 31-MAY-2000 (first entry)
XX DE Rat sensory transduction G-protein coupled receptor-B3 variant #1.
XX KW Rat; sensory transduction; G-protein coupled receptor-B3; GPCR-B3;
XX KW sensory cell; taste receptor cell; screen; taste modulator;
XX KW pharmaceutical; food industry; taste topographic map; tongue; variant.
XX OS Rattus sp.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Misc-difference 33 /note= "Wild type Leu substituted by Ile"
XX FT WO200006592-A1.
XX PD 10-FEB-2000.
XX PF 27-JUL-1999; 99WO-US17099.
XX PR 28-JUL-1998; 98US-0094465.
XX PA (REGC) UNIV CALIFORNIA.
XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX PI Zuker CS, Adler JE, Lindemeier J, Ryba N, Hoon M;
XX WPI; 2000-205451/18.
XX DE New isolated sensory transduction G-protein coupled receptor, useful
XX PT for developing products for use in studying and modulating the taste
XX PT transduction pathway -
XX PS Disclosure; Page -; 83pp; English.
XX CC The present sequence is a polymorphic variant #1 of rat taste cell
XX CC specific G-protein coupled receptor, GPCR-B3 which is involved in
XX CC sensory transduction. GPCR-B3 is specifically expressed in foliate
XX CC and fungiform cells, with lower expression in circumvallate taste
XX CC receptor cells of the tongue. The protein has an extracellular domain,
XX CC seven transmembrane domains and an intracellular domain.
XX CC The GPCR-B3 sequence is used to screen compounds that modulate sensory
XX CC signalling in taste cells, especially taste modulators useful in
XX CC pharmaceutical and food industries to customise taste. The sequence
XX CC can also be used as probe for identifying taste cells and
XX CC subsets of taste receptor cells such as foliate, fungiform and
XX CC circumvallate. Such probes are also useful to generate taste
XX CC topographic maps that elucidate the relationship between the taste
XX CC cells of the tongue and sensory neurons leading to taste centres
XX CC in the brain.
XX CC Note: The present sequence is not given in the specification but is
XX CC derived from rat GPCR-B3 sequence shown in page 75 (AAY45021).

Search completed: November 1, 2001, 10:14:22
Job time: 3715 sec

04 Wild type Glu substituted by Asp"

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:13:25 ; Search time 26.44 Seconds
(without alignments)
466.476 Million cell updates/sec

Title: US-09-775-181-4
Perfect score: 3199
Sequence: 1 MGAMAYPLLCLLLAQLGLG.....YMTAVGMWSLVSYDGLTIFQ 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133.5	4.2	1079	1 US-08-485-588-8	Sequence 8, Appli
2	133.5	4.2	1079	1 US-08-484-565-8	Sequence 8, Appli
3	133.5	4.2	1079	2 US-08-480-751-8	Sequence 8, Appli
4	133.5	4.2	1079	2 US-08-943-986-8	Sequence 8, Appli
5	133.5	4.2	1079	3 US-08-353-784-8	Sequence 8, Appli
6	133.5	4.2	1079	3 US-08-484-719B-8	Sequence 8, Appli
7	127	4.0	1078	1 US-08-485-588-7	Sequence 7, Appli
8	127	4.0	1078	1 US-08-484-565-7	Sequence 7, Appli
9	127	4.0	1078	2 US-08-480-751-7	Sequence 7, Appli
10	127	4.0	1078	2 US-08-943-986-7	Sequence 7, Appli
11	127	4.0	1078	3 US-08-353-784-7	Sequence 7, Appli
12	127	4.0	1078	3 US-08-484-719B-7	Sequence 7, Appli
13	123	3.8	1085	1 US-08-485-588-5	Sequence 5, Appli
14	123	3.8	1085	1 US-08-484-565-5	Sequence 5, Appli
15	123	3.8	1085	2 US-08-480-751-5	Sequence 5, Appli
16	123	3.8	1085	2 US-08-943-986-5	Sequence 5, Appli
17	123	3.8	1085	3 US-08-353-784-5	Sequence 5, Appli
18	123	3.8	1085	3 US-08-484-719B-5	Sequence 5, Appli
19	121.5	3.8	1059	4 US-09-134-513-2	Sequence 2, Appli
20	119.5	3.7	1088	1 US-08-485-588-6	Sequence 6, Appli
21	119.5	3.7	1088	1 US-08-484-565-6	Sequence 6, Appli
22	119.5	3.7	1088	2 US-08-480-751-6	Sequence 6, Appli
23	119.5	3.7	1088	2 US-08-943-986-6	Sequence 6, Appli
24	119.5	3.7	1088	3 US-08-353-784-6	Sequence 6, Appli
25	119.5	3.7	1088	3 US-08-484-719B-6	Sequence 6, Appli
26	118.5	3.7	1058	2 US-08-687-289A-5	Sequence 5, Appli
27	118	3.7	872	3 US-08-337-797A-2	Sequence 2, Appli

28 118 3.7 872 3 US-09-258-523-2 Sequence 2, Appli
29 116 3.6 906 1 US-08-486-270-2 Sequence 2, Appli
30 116 3.6 906 3 US-08-367-264-2 Sequence 2, Appli
31 116 3.6 1056 2 US-08-687-289A-8 Sequence 8, Appli
32 115 3.6 906 5 PCT-US91-09422-17 Sequence 17, Appli
33 115 3.6 1199 1 US-08-041-538-2 Sequence 2, Appli
34 115 3.6 1199 1 US-08-463-642-2 Sequence 2, Appli
35 115 3.6 1199 1 US-08-465-602-2 Sequence 2, Appli
36 115 3.6 1199 2 US-08-465-157-2 Sequence 2, Appli
37 115 3.6 1199 5 PCT-US91-09422-2 Sequence 2, Appli
38 114.5 3.6 879 1 US-08-072-574-6 Sequence 6, Appli
39 114.5 3.6 879 1 US-08-486-270-6 Sequence 6, Appli
40 114.5 3.6 879 3 US-08-367-264-6 Sequence 6, Appli
41 114.5 3.6 1219 2 US-08-687-289A-6 Sequence 6, Appli
42 114 3.6 1180 4 US-08-660-148-2 Sequence 2, Appli
43 114 3.6 1212 4 US-08-660-148-5 Sequence 5, Appli
44 113 3.5 877 1 US-08-072-574-12 Sequence 12, Appli
45 113 3.5 877 1 US-08-486-270-12 Sequence 12, Appli

ALIGNMENTS

RESULT 1
US-08-485-588-8
; Sequence 8, Application US/08485588
; Patent No. 5688938
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: Forrest H. Fuller
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,588
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 9
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451

;; FILING DATE: 23 August, 1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Heber, Sheldon O.
;; REGISTRATION NUMBER: 38,179
;; REFERENCE/DOCKET NUMBER: 213/005
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 489-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1079 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-485-588-8

Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;
QY 235 HRRGNQGRGLGHSWRKDGKSHFKWSPYLECENGSKYKPGWLVTLSAIYGLQP 294
DB 377 HEEG---GNRLNSSTAPRPLCTGDNINSVETPYMDYEHRLRISYNYLAVYSIAHALQD 433
QY 295 --NLVPERGVKVDINLQKVDIDQCSSDGSFSGTHKHLNNS-----ECMPIKG 342
DB 434 IYCLPG-RGLF---TNGSCADIKV--EAMQVLKHLRHLNFTNMGGQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNRRRGGPDQHS----- 379
DB 488 NYSIINHLSPEDGSIVFKEVGYNYVAKKGERLFINEEKILWGSFREVFPNSCRDQ 547
QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYL 417
DB 548 AGTRKGIIEGPTCCFCEVCEPDGYSGETDASACDKCPDFFWSNENHTSCIAKEIEFLA 607
QY 418 -----LAIISFQGLCMLLDFVSMVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471
DB 608 WTEPFGIALTFAVLGIFLTAFLGVFKFRNTPIVKATNRELSYLLFLSLCCFSSSLF 667
QY 472 LYFEPSTERCILLRWALLGATVYGTTLKHLRVKLVFLSRTAQRIPYMTGGRVMRLA 531
DB 668 FIFGPDWTCRLQPAFGISFVLCSICILVKTNRVLLVFEAKIPTSFHRKWWGLNQLFLL 727
QY 532 V-----ILLVWFELIGWTSVYCONLEKQISLIGOGKTSDDLIFNMCLIDRWDMYMTAV 584
DB 728 VFLCTFMQLICIWLITAPPSSY-RNHELE-----DELIIFTTC---HEGSLMAL 773
QY 585 GMSLVSY 592
DB 774 G-SLIGY 779

RESULT 2
US-08-484-565-8
; Sequence 8, Application US/08484565
; Patent No. 5763569
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California

COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1079 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-565-8

Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;
QY 235 HRRGNQGRGLGHSWRKDGKSHFKWSPYLECENGSKYKPGWLVTLSAIYGLQP 294
DB 377 HEEG---GNRLNSSTAPRPLCTGDNINSVETPYMDYEHRLRISYNYLAVYSIAHALQD 433
QY 295 --NLVPERGVKVDINLQKVDIDQCSSDGSFSGTHKHLNNS-----ECMPIKG 342
DB 434 IYCLPG-RGLF---TNGSCADIKV--EAMQVLKHLRHLNFTNMGGQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNRRRGGPDQHS----- 379
DB 488 NYSIINHLSPEDGSIVFKEVGYNYVAKKGERLFINEEKILWGSFREVFPNSCRDQ 547
QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYL 417
DB 548 AGTRKGIIEGPTCCFCEVCEPDGYSGETDASACDKCPDFFWSNENHTSCIAKEIEFLA 607
QY 418 -----LAIISFQGLCMLLDFVSMVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471
DB 608 WTEPFGIALTFAVLGIFLTAFLGVFKFRNTPIVKATNRELSYLLFLSLCCFSSSLF 667

Db 774 G--SLIGY 779

||: |

RESULT 5

US-08-353-784-8

; Sequence 8, Application US/08353784

; Patent No. 6011068

; GENERAL INFORMATION:

; APPLICANT: Edward F. Nemeth, Edward M.

; APPLICANT: Brown, Steven C. Hebert,

; APPLICANT: Bradford C. Van Wagonen, Manuel

; APPLICANT: F. Balandrin, Forrest H. Fuller,

; APPLICANT: Eric G. DelMar, and Scott T. Moe

; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

; TITLE OF INVENTION: MOLECULES

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: First Interstate World Center

; STREET: Suite 4700

; STREET: 633 West Fifth Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: FASTSEQ

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/353,784

; FILING DATE: 9 December, 1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA: including application

; PRIOR APPLICATION DATA: described below: 8

; APPLICATION NUMBER: PCT/US/94/12117

; FILING DATE: 21 October, 1994

; APPLICATION NUMBER: U.S. 08/292,827

; FILING DATE: 23 August, 1994

; APPLICATION NUMBER: U.S. 08/141,248

; FILING DATE: 22 October, 1993

; APPLICATION NUMBER: U.S. 08/009,399

; FILING DATE: 23 February, 1993

; APPLICATION NUMBER: U.S. 08/017,127

; FILING DATE: 12 February, 1993

; APPLICATION NUMBER: U.S. 07/934,161

; FILING DATE: 21 August, 1992

; APPLICATION NUMBER: U.S. 07/834,044

; FILING DATE: 11 February, 1992

; APPLICATION NUMBER: U.S. 07/749,451

; FILING DATE: 23 August, 1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Heber, Sheldon O.

; REGISTRATION NUMBER: 38,179

; REFERENCE/DOCKET NUMBER: 209/069

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1079 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-353-784-8

Query Match 4.2%; Score 133.5; DB 3; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;

Matches	93;	Conservative	58;	Mismatches	182;	Indels	95;	Gaps	18;
---------	-----	--------------	-----	------------	------	--------	-----	------	-----

Qy	235	HRCPNPGRGLGHSWRKDGKDSHKFKWSPYLECENGSYKPGWLVTLSAIYGLQP	294
Db	377	HEEG---GNBLNLSSTAFRLCTGDEININSVETPYMDYEHRLISYVNYLAVYSIAHALQD	433
Qy	295	--NLVPFRCRGMKVDINLQKVDIDOCSSDGFSGTHKCHLNS-----ECMPIKG	342
Db	434	IYTCLPG-RGLF---TNGSCADIKV--EAWQVLKHLRLHFNNTNMNGEQVTFDECGDING	487
Qy	343	LGFLV-----GAYECICRAG-----FYHPGVLVNNFRFRGPOQHS-----	379
Db	488	NYSIINMHLSPEDGSIYFKVEGVYNYVAKKGERLFINEEKILWSGFSREVPFNSCRDQC	547
Qy	380	-GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYL	417
Db	548	AGTRKGIIIEPTCCFCEVCEPGEYSGETDASACDKCPDDFSWNNHNSCTAKEFEFLA	607
Qy	418	-----LAIISFOGLCMLLDFVSMLVVYHFRKAKSIASGLIILETFILFCSLLLYPFWVI	471
Qy	608	WTEPFGIALTFAVLGIFLTAFLVGVIKFRNTPIVKATNRELSYLLLSLCCCFSSLF	667
Qy	472	LYEPSTFCILLRWARLLGFATVYGTVTLKLRVLKVFLSRTAQRIPYWTGGRVWRMLA	531
Db	668	FIGEPQDWTCLROPAGFISFVILCISCLVKTNRVLLVFEAKIPTSFHRKWMGLNLOFLL	727
Qy	532	V-----ILLVFWFLIGHTWSVCQNLKQISLIGOGKTSDLHINMCLIDRWYMTAV	584
Db	728	VLCTFMQILICIWLYTAPSSY-RNHELE-----DEIIFTC---HEGSLMAL	773
Qy	585	GMWSVSY	592
Db	774	G--SLIGY	779

RESULT 6

US-08-484-719B-8
; Sequence 8, Application US/08484719B
; Patent No. 6031003
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth, Edward M.
; APPLICANT: Brown, Steven C. Hebert,
; APPLICANT: Bradford C. Van Wagenen,
; APPLICANT: Manuel F. Balandrin,
; APPLICANT: Forrest H. Fuller, Eric G.
; APPLICANT: Delmar, Scott T. Moe
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS Word
; SOFTWARE: FastSeq for Windows Version 3.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,719B
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827

? FILING DATE: 23 August, 1994
? APPLICATION NUMBER: U.S. 08/141,248
? FILING DATE: 22 October, 1993
? APPLICATION NUMBER: U.S. 08/009,389
? FILING DATE: 23 February, 1993
? APPLICATION NUMBER: U.S. 08/017,127
? FILING DATE: 12 February, 1993
? APPLICATION NUMBER: U.S. 07/934,161
? FILING DATE: 21 August, 1992
? APPLICATION NUMBER: U.S. 07/834,044
? FILING DATE: 11 February, 1992
? APPLICATION NUMBER: U.S. 07/749,451
? FILING DATE: 23 August, 1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Douglas C. Murdock
? REGISTRATION NUMBER: 37,549
? REFERENCE/DOCKET NUMBER: 213/007
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (213) 489-1600
? TELEFAX: (213) 955-0440
? TELEX: 67-3510
? INFORMATION FOR SEQ ID NO: 8:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1079 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-484-719B-8

Query Match	Score	DB 3	Length
4.28	133.5	1079	

Query Match 4.2%; Score 133.5; DB 3; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.00041;
Matches 93; Conservative 58; Mismatches 182;
Indels 95; Gaps 18;

Qy	235	HRRPNQOPRGLGSHWRKRGDGLGDKSHFKWSPPYLECENGSKYKPGWLVTLSAIYGLQ	294
Db	377	HEEG---GNRLNNSAFRPLCTGDEINSVETPYMDYEHRLRISYVNYLAVYSIAHALQD	433
Qy	295	--NLVPEFRGMVKVDINLQKVDIDOCSSDGWFSGTIKCHLNN--	342
Db	434	IYTCLPG-RGLF---TNGSCADIKV--EAKQVLKHLRLHNTNNMGEQVTFECCDGLVG	487
Qy	343	LGFWL-----GAYECTCKAG--FYHPGVLVNNFRRRGPQOHIS----	379
Qy	488	NYSIINWHLSPEDGSIVFEKGVYNYVAKKGERLFINEEKILWSGFSREVFPSCNRDQC	547
Qy	380	-GSTKDVSEB---AYVCLPCREG-----CPCADD-----SPCFVQEDOKYLR	417
Db	548	AGTRKGIIEGPTCCFCEVCPEGSEGTDSACOKCPDQFWSNHNHTSIKAELEFLA	607
Qy	418	-----LAITSFQGLMLLDFVSNLVVYHFRKAKSIASGLIILETLFGSLLLYPPVVI	471
Db	608	WTEPFEGIALFLFVGLGFLTAFLVGLFKFRNTPIVKATNRELSQLLLLSLCCFSSSLF	667
Qy	472	LYFEPSTERCILLRWARELGFATVGTUULKHUVLKVFILSTAQRIPTYMTGGRVNRMLA	531
Db	568	FIGEPQDWTCLRQAFGIFVLCISCLVKTNRVLVYFEAKIPTSFHRKRWGSLNQFL	727
Qy	532	V-----ILLVFWFLIGWTSVVCNLEKQISLIGQKTSDDLHINMCLIDRWDMYTA	584
Db	728	VFLCTFQWILICIWLTPAPSSY-RNHELE-----DELIIFTC---HEGSJMAL	773
Qy	585	GMWSLVSY	592
Db	774	G--SLIG	779

RESULT 7

US-08-485-588-7
; Sequence 7, Application US/08485588
; Patent No. 5688938
; GENERAL INFORMATION:

APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,588
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/005
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-588-7

Query Match 4.0%; Score 127; DB 1; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;
Qy 311 QKVIDQSSDGFSGTHKC---HUNNSE-CMPKGLGVLGAYECICKAG---FYHPGV 363
Db 475 EQVTEDEC---GDLVGNYSINWHLSPEDGSIVFKEVGY---YNYAKKGERLFINEEK 527
Qy 364 LPVNNFRFRGPDQH-----ISGSKVDVSEE-----AYVCLPCREG-----CPFCAD 404

Db 528 ILWSGFSREVPFNSCRDLACTRKGIIETGPTCFECVECPDGEYSDETASACNKC PD 587
Qy 405 D-----SPCFVQEDKYLR-----LAIISFOGLCMLLDFVSMVWYHFRKAKSIRASG 451
Db 588 DFWSNENHTSCIAKEIEFLSWTEPGIALTFVAVLGIFLTAFLVGLGVFKFRNTPIVKATN 647
Qy 452 LILLETLFGSLILYFPVVILYFEPSTFCRILLRWARLLGFATVYGTVTLKLRHVKVFL 511
Db 648 RELSYLLLFSLCCPSSSLFFIGEPQDWTCLRQPAFGISFVLCISCLVKTNRVLLVFE 707
Qy 512 SRTAQRIPYMTGGRVNRMLAV-----ILLVVFELGWTSSV-CQMLEKQISLI----- 559
Db 708 AKIPTSFHRKMWMLNLQFLVFLCTFMQIVTCVIMVLTAPPSSYRNQLEDEIIFITCHE 767
Qy 560 GQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLGYTCLL 783
RESULT 8
US-08-484-565-7
Sequence 7, Application US/08484565
Patent No. 5763569
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006

TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-565-7

Query Match 4.0%; Score 127; DB 1; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVDIDOCSSDGNFSGTHKC---HLNNS-CMPKIGLGVLGAYECICKAG---FYHPGV 363
475 EQVTFDEC---GDLVGNYSIINHLSPEDESGSIVFKEVGY---YNYAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSTKDYSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVPFNSCRDCLAGTRKGIIEGPTCCPEVCPEDEYSDETDASACNKKPD 587
QY 405 D-----SPCFVQEDKYLK-----LAIISFOGLCMLLDFVSMVYVYHFRKAKSIRASG 451
588 DFWSNENHTSCIAKEIEFLSWTEPFEGIALTLFAVLGIFLTAFLGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYFPVILYFPEPSTFRICLLRWALLGFAIVYGVVTKLHRLVKVFL 511
648 RELSYLLFLSLLCCFSSSLFFIGEPQDWTCLRPQAFGISFVLCISCLVKTNRVLLVFE 707
QY 512 SRTAQRIPYMTGGVRMRLAV-----ILLVWFELIGHTSSV-CQLEKQISLI----- 559
708 AKIPTSFRKRWGMLNQFLVFLCTFMQIVICVILWLTAPPSSYRNOELEDIEIIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
768 GSLMALGFLIGYTCLL 783

RESULT 9
US-08-480-751-7
Sequence 7, Application us/08480751
Patent No. 5858684
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,751
FILING DATE: 7 June, 1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-751-7

Query Match 4.0%; Score 127; DB 2; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVDIDOCSSDGNFSGTHKC---HLNNS-CMPKIGLGVLGAYECICKAG---FYHPGV 363
475 EQVTFDEC---GDLVGNYSIINHLSPEDESGSIVFKEVGY---YNYAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSTKDYSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVPFNSCRDCLAGTRKGIIEGPTCCPEVCPEDEYSDETDASACNKKPD 587
QY 405 D-----SPCFVQEDKYLK-----LAIISFOGLCMLLDFVSMVYVYHFRKAKSIRASG 451
588 DFWSNENHTSCIAKEIEFLSWTEPFEGIALTLFAVLGIFLTAFLGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYFPVILYFPEPSTFRICLLRWALLGFAIVYGVVTKLHRLVKVFL 511
648 RELSYLLFLSLLCCFSSSLFFIGEPQDWTCLRPQAFGISFVLCISCLVKTNRVLLVFE 707
QY 512 SRTAQRIPYMTGGVRMRLAV-----ILLVWFELIGHTSSV-CQLEKQISLI----- 559
708 AKIPTSFRKRWGMLNQFLVFLCTFMQIVICVILWLTAPPSSYRNOELEDIEIIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
768 GSLMALGFLIGYTCLL 783
RESULT 10
US-08-943-986-7
Sequence 7, Application US/08943986
Patent No. 5962314
GENERAL INFORMATION:
APPLICANT: Edward M. Brown

APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,986
FILING DATE: 03-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,565
FILING DATE: 7-June-1995
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-986-7

Query Match 4.0%; Score 127; DB 2; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;
Qy 311 OKVIDOCSSDGSFSTHCK---HLNNE-CMPKGLGFLVGLAYECICKAG---FYHFGV 363
Db 475 EQVTFEC---GDLVGNYSIINHLSPDGSIVFKEVY-----YNVYAKGERLFINEEK 527
Qy 364 LPVNNFRRRGPDQH-----ISGSKDVSEE-----AYVCLPREG-----CPFCAD 404
Db 528 ILWSGFSREVPFNSCRDLAGTRKGIIEGPTCCFECVCEPCDGEYSDETASACNCPD 587

Qy 405 D-----SPOFVQEDKYLR-----LAIISQGLCMLLDFFVSMVLVYVYHFRKAKSIRASG 451
Db 588 DFWSNENHSTCIAKEIEFLSWTEPEFGIALTLFAVLGIELTFAVLGFIKFRNTPIVKATN 647
Qy 452 LILLETLFGSLLYFPVILYFEPSTERCILLRWALLGFATVYGTVLKLRHVKVFL 511
Db 648 RELSYLLFLSLCCFSSSLFFIGEPQDWTCLRQPAFGISFVLCISCLVKTNRVLLVFE 707
Qy 512 SRTAQRIPYMTGGRVMRLAV-----ILLVWFELIGWTSV-CQNEKQISLI----- 559
Db 708 AKIPTSFHRKWWGLNLQFLVFLCTFMQIVICVIMLYTAPSSYRNQOELEDEIFITCHE 767
Qy 560 CQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLIGVTCLL 783
RESULT 11
US-08-353-784-7
Sequence 7, Application US/08353784
Patent No. 6011068
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert,
APPLICANT: Bradford C. Van Wagenen, Manuel
APPLICANT: F. Balandrin, Forrest H. Fuller,
APPLICANT: Eric G. DelMar, and Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,784
FILING DATE: 9 December, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 8
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 209/069
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-784-7

Query Match 4.0%; Score 127; DB 3; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVIDQCSSDGFSGTHKC---HLNNS-CMPIKGLGVLGAYECICKAG---FYHPGV 363
Db 475 EQVTFDEC---GDLVGNYSIINHLSPEDEGSIYFKEVGY---YNYVAKKGERLFINEEK 527
364 LPVNNFRRRGPDOH-----ISGSTKDVSEE---AYVCLPCREG-----CPFCAD 404
528 ILWSGFSREVFPFNSCRDCLAGTRKGIIEGPTCCFCEVCEPDGEYSDETDASACNKKCPD 587
QY 405 D-----SPCFVOEDKYLK-----LAISFOGLCMLDFVSMVYHFRKAKSIASG 451
Db 588 DFWSNENHTSCIAKEIEFLSWTEPFGIALTLFAVLGIFLTAFLVGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYPVVTLYPEPSTFRICILLRWARLLGFAIVGTIVLKLHRLVKVFL 511
Db 648 RELSYLLFLSLCCFSSSFFIGEPQDWTCLRPQAFGISFVLCISILYKTRNVLVFE 707
QY 512 SRTAQRIPTYMTGVRMRLAV-----ILLVWFVFLIGWTSV-CONLEKQISLI----- 559
Db 708 AKIPTSFRKRWGGLNQFLVFLCTFMQIVICVILWLYTAPPSSYRNOELEDIEIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLIGYTCLL 783

RESULT 12
US-08-484-719B-7
Sequence 7, Application US/08484719B
Patent No. 6031003
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert,
APPLICANT: Bradford C. Van Wagenen,
APPLICANT: Manuel F. Balandrin,
APPLICANT: Forrest H. Fuller, Eric G.
APPLICANT: Delmar, Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS Word
SOFTWARE: FastSeq for Windows Version 3.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,719B
FILING DATE: 7 June, 1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Douglas C. Murdock
REGISTRATION NUMBER: 37,549
REFERENCE/DOCKET NUMBER: 213/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-719B-7

Query Match 4.0%; Score 127; DB 3; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0018;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVIDQCSSDGFSGTHKC---HLNNS-CMPIKGLGVLGAYECICKAG---FYHPGV 363
Db 475 EQVTFDEC---GDLVGNYSIINHLSPEDEGSIYFKEVGY---YNYVAKKGERLFINEEK 527
QY 364 LPVNNFRRRGPDOH-----ISGSTKDVSEE---AYVCLPCREG-----CPFCAD 404
Db 528 ILWSGFSREVFPFNSCRDCLAGTRKGIIEGPTCCFCEVCEPDGEYSDETDASACNKKCPD 587
QY 405 D-----SPCFVOEDKYLK-----LAISFOGLCMLDFVSMVYHFRKAKSIASG 451
Db 588 DFWSNENHTSCIAKEIEFLSWTEPFGIALTLFAVLGIFLTAFLVGVFKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLYPVVTLYPEPSTFRICILLRWARLLGFAIVGTIVLKLHRLVKVFL 511
Db 648 RELSYLLFLSLCCFSSSFFIGEPQDWTCLRPQAFGISFVLCISILYKTRNVLVFE 707
QY 512 SRTAQRIPTYMTGVRMRLAV-----ILLVWFVFLIGWTSV-CONLEKQISLI----- 559
Db 708 AKIPTSFRKRWGGLNQFLVFLCTFMQIVICVILWLYTAPPSSYRNOELEDIEIFITCHE 767
QY 560 GQKTSDDLIFNMCLI 575
Db 768 GSLMALGFLIGYTCLL 783

RESULT 13
US-08-485-588-5
Sequence 5, Application US/08485588
Patent No. 5688938
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: Forrest H. Fuller

APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
SUITE: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Hebert, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/005
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1085 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-588-5

Query Match 3.8%; Score 123; DB 1; Length 1085;
Best Local Similarity 21.1%; Pred. No. 0.0046;
Matches 64; Conservative 51; Mismatches 135; Indels 54; Gaps 12;
Qy 311 QKVIDQSSDCWFSGTHKC---HLNNS-CMPKGLGVLGAYECICKAG---FYHPGV 363
Db 476 EQVTFDEC---GDLAGNYSIINWHLSPDGSIVPKVGY----YNYAKKGERFINDEK 528
Qy 364 LPVNNFRRRGPDQH-----ISGSTKDVSE-----AYVCLPREG-----CPFCAD 404
Db 529 ILWGSFSEVPFNSCRLAGTRKGIETEPTCCFCEVCEPDGYESDSDASACDKCPD 588
Qy 405 D-----SPCFVQEDKYLK-----LAIISFQGLCMLLDVFSMLVYVHFRKAKSIRASG 451

Db 589 DFWSNENHTSCIAKEIEFISWTEPGIALTLFAVLGIFLTAFLVGLFKFRNTPIVKATN 648
Qy 452 LILLETILFGSLLLYFPVYLVFEPSTERCILLRWALLGFATVYGTVLKLRVLYKVPFL 511
Db 649 RELSYLLLSLCCFSSSLFFIGEPQDWTCTLRQPAFGISFVLCISCLVKTNRVLLVFE 708
Qy 512 SRTAQRIPTYMTGGRVYRMLAVILLVVFVFLIG--WTSSVCQNLKQIQLIGQGTSDHLI 569
Db 709 AKIPTSFHRKWWGLNLQFLLVFLCTFMQIVICAIWLTAPPSYRNHEL-----EDEII 762
Qy 570 FNMCM 573
Db 763 FITC 766
RESULT 14
US-08-484-565-5
Sequence 5, Application US/08484565
Patent No. 5763569
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
SUITE: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Hebert, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440

OM of: US-09-775-181-4 to: GenEmbl.* out_format : pfs

Date: Nov 3, 2001 1:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame*_p2n.model -DEV=xlh
-Q/cgnt2/_USPTO_spool/US09775181/runat_30102001_161056_13810/app_query.fasta_1.1955
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOPC=0.000 -LOPEXT=0.000 -OGAPOP=4.500
-OGAEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09775181@cgnl_1.11947 -NCPU=6
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
-WAIT -THREADS=1

arch information block:

Query: US-09-775-181-4
Query length: 599
Database: GenEmbl.*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 4100.330000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_hgt21:AL358932	+	1607.00	1880.63	1.6e-96	11117	! AL358932 Homo sapiens chrom
gb_hgt20:AL139821	+	1607.00	1875.72	3.1e-96	199526	! AL139821 Homo sapiens chrom
gb_hgt20:AL355587	+	489.00	540.44	7.3e-22	151218	! AL355587 Homo sapiens chrom
gb_hgt21:AL359974	-	409.00	445.14	1.5e-16	143878	! AL359974 Homo sapiens chrom
gb_hgt20:AL161654	-	403.00	437.09	4.1e-16	159596	! AL161654 Homo sapiens chrom
gb_hgt5:AC017271	-	259.50	287.92	8.4e-08	10892	! AC017271 Drosophila melanoga
gb_hgt1:AE009845	-	259.50	264.22	1.8e-06	183766	! AE009845 Drosophila melanoga
gb_hgt1:AE003575	+	259.50	259.39	3.3e-06	327209	! AE003575 Drosophila melanoga
gb_hgt20:AL354976	+	202.00	195.28	0.0122	186964	! AL354976 Homo sapiens chrom
gb_hgt22:AL445244	-	200.00	194.18	0.0140	160390	! AL445244 Homo sapiens chrom
gb_hgt4:AC014084	-	183.50	174.99	0.1646	150133	! AC014084 Drosophila melanoga
gb_hgt1:AC007772	-	183.50	173.71	0.2079	186607	! AC007772 Drosophila melanoga
gb_hgt1:AC007772	-	183.50	172.17	0.2205	197071	! AC007772 Drosophila melanoga
gb_hgt2:AE003732	-	183.50	171.26	0.2657	234369	! AE003732 Drosophila melanoga
gb_hgt1:DMX99675	+	164.00	183.61	0.0845	3325	! X99675 D.melanogaster mRNA fol
gb_hgt1:HSPCAR1	+	161.00	180.25	0.0838	3234	! X81086 H. sapiens Pcar1 gene.
gb_hgt1:S83176	+	161.00	180.25	0.0838	3234	! S83176 Homo sapiens calcium-se
gb_hgt10:HMCASR	+	161.00	179.93	0.0873	3361	! D50855 Human mRNA for Ca-sensi
gb_hgt10:HSU20759	+	161.00	178.94	0.0992	3783	! U20759 Human parathyroid cell
gb_hgt1:AR012624	+	161.00	178.88	0.0999	3809	! AR012624 Sequence 3 from paten
gb_hgt1:AR028467	+	161.00	178.88	0.0999	3809	! AR028467 Sequence 3 from paten
gb_hgt1:AR078219	+	161.00	178.88	0.0999	3809	! AR078219 Sequence 3 from paten
gb_hgt2:175053	+	161.00	178.88	0.0999	3809	! I75053 Sequence 3 from paten
gb_hgt4:AC014500	+	160.00	163.45	0.7228	20791	! AC014500 Drosophila melanoga
gb_hgt1:AC009257	+	160.00	143.72	9.07	218565	! AC009257 Drosophila melanoga
gb_hgt1:AE003461	+	160.00	141.11	12.69	298616	! AE003461 Drosophila melanoga
gb_hgt1:S67307	+	152.50	166.02	0.5199	5249	! S67307 Ca(2+)-sensing receptor
gb_hgt1:AR012622	+	152.50	165.98	0.5227	5275	! AR012622 Sequence 1 from paten
gb_hgt1:AR028465	+	152.50	165.98	0.5227	5275	! AR028465 Sequence 1 from paten
gb_hgt1:AR078217	+	152.50	165.98	0.5227	5275	! AR078217 Sequence 1 from paten
gb_hgt2:175051	+	152.50	165.98	0.5227	5275	! I75051 Sequence 1 from paten
gb_hgt1:AR012623	+	150.00	163.43	0.7251	5006	! AR012623 Sequence 2 from paten
gb_hgt1:AR028466	+	150.00	163.43	0.7251	5006	! AR028466 Sequence 2 from paten
gb_hgt1:AR078218	+	150.00	163.43	0.7251	5006	! AR078218 Sequence 2 from paten
gb_hgt2:175052	+	150.00	163.43	0.7251	5006	! I75052 Sequence 2 from paten
gb_hgt1:AR028465	+	150.00	163.42	0.7255	5009	! G28586 Human STS SHCC-35350, s
gb_hgt10:HSU20760	+	150.00	163.42	0.7255	5009	! U20760 Human extracellular cal
gb_hgt1:AR084671	+	148.50	163.51	0.7171	4000	! AR084671 Sequence 2 from paten
gb_hgt4:AC014621	-	148.00	150.23	3.94	18160	! AC014621 Drosophila melanoga
gb_hgt2:AE003846	-	148.00	126.75	79.99	298469	! AE003846 Drosophila melanoga

gb_roi:AF110178 + 145.50 158.84 1.31 4550 ! AF110178 Mus musculus calci
gb_roi:AF128842 - 144.50 159.68 1.17 3569 ! AF128842 Mus musculus extra
gb_in3:LMFL6520 - 144.50 139.33 15.95 40438 ! AL446005 Leishmania major
gb_ov:AB008857 + 144.00 159.00 1.28 3606 ! AB008857 Fugu rubripes gene
gb_in2:AF318274 + 144.00 158.30 1.40 3918 ! AF318274 Drosophila melanog

seq_name: gb_hgt21:AL358932

seq_documentation_block:

LOCUS AL358932 111117 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.

ACCESSION AL358932

VERSION AL358932.3 GI:9930957

KEYWORDS HTG; HTGS_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE Sims,S.

AUTHORS Direct Submission

TITLE

JOURNAL

Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Aug 27, 2000 this sequence version replaced gi:9926667.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA395P8

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 90261 bases at least Q40

Consensus quality: 99226 bases at least Q30

Consensus quality: 104490 bases at least Q20

Insert size: 108717; sum-of-contigs

Insert size: 178772; 24.0% error; agarose-fp

Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality

coverage: 1.42x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently

consists of 25 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 2287: contig of 2287 bp in length

* 2288 2387: gap of 100 bp

* 2388 6317: contig of 3930 bp in length

* 6318 6417: gap of 100 bp

* 6418 9725: contig of 3308 bp in length

* 9726 9825: gap of 100 bp

* 9826 16607: contig of 6782 bp in length

* 16608 16707: gap of 100 bp

* 16708 20159: contig of 3452 bp in length

* 20160 20259: gap of 100 bp

* 20260 23641: contig of 3382 bp in length

* 23642 23741: gap of 100 bp

* 23742 27399: contig of 3658 bp in length

* 27400 27499: gap of 100 bp

* 27500 32754: contig of 5255 bp in length

* 32755 32854: gap of 100 bp

* 32855 35240: contig of 2386 bp in length

* 35241 35340: gap of 100 bp

* 35341 40948: contig of 5608 bp in length

* 40949 41048: gap of 100 bp


```
83853 CAGGACGACCTGGATTGGTACCAGCGCTGGTGTGGAGCCCTCTGGAGGG 83902
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeus 184
83903 CGAGCCAGCATCTCCCGGGGGCCATCACCTTCAGCACCGATTGCGTGT 83952
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
83953 CGGCACCGGGCCCCACAGGTCTTCCTCCAGGCGCACGCGGAGGAGCGCG 84002
201 ileLeuLeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaTh 217
84003 ATCTCTGCTCCAGACCTGTCTCTCTCCGACCCCACTTGGCCAGGCCAC 84052
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
84053 TCTGGAGACCGAGTGTCTCCACGGCTCTCCGGCGCAAGTGGAGGCCCACT 84102
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
84103 TACACCGCGCGGGCCCCCAATCAGGGGGCCCCGGGGCTGGGCCACAGCTGG 84152
251 ArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSerPr 267
84153 CGGCGCAAGGAGCGGCTCGCGGGGACAGAGCCACTTCAAGTGTCTCC 84202
267 oProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
84203 GCCTTATCTGGAGTCCGAGAACGGGAGTTACAAGCCCGGGTGGTGGTTA 84252
284 hrLeuSerSerAlaIleIleThrGlyLeuGlnProAsnLeuValProGluPhe 300
84253 CTCCTTCTCTGCCATCTACCGGTTTGCAGCCTTAACCTGGTCCCGGAATC 84302
301 Arg 301
84303 AGG 84305
seq_name: gb_btg20:AL139821
seq_documentation_block:
LOCUS AL139821 199526 bp DNA HTG 17-MAR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-80K21, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL139821
VERSION AL139821.7 GI:13396382
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Plumb,B.
Direct Submission
Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:10185474.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA80K21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big dye; 100% of reads
Consensus quality: 197300 bases at least Q40
Consensus quality: 198299 bases at least Q30
Consensus quality: 198676 bases at least Q20
Insert size: 199026; sum-of-contigs
```

```
Insert size: 144047; 11.6% error; agarose-fp
Quality coverage: 7.42x in Q20 bases; sum-of-contigs Quality
coverage: 10.32x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 37012: contig of 37012 bp in length
* 37013 37112: gap of 100 bp
* 37113 163560: contig of 126448 bp in length
* 163561 163660: gap of 100 bp
* 163661 171563: contig of 7903 bp in length
* 171564 171663: gap of 100 bp
* 171664 186786: contig of 15123 bp in length
* 186787 186886: gap of 100 bp
* 186887 196057: contig of 9171 bp in length
* 196058 196157: gap of 100 bp
* 196158 199526: contig of 3369 bp in length.
```

FEATURES

source

```
1..199526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RP11-80K21"
/clone_lib="RPC1-11.1"
1..37012
/note="assembly_fragment:00668"
fragment_chain:1
clone_end:T7
vector_side:left"
37113..163560
/note="assembly_fragment:02214"
fragment_chain:1
clone_end:SP6
vector_side:right"
163661..171563
/note="assembly_fragment:00510"
171664..186786
/note="assembly_fragment:01538"
186887..196057
/note="assembly_fragment:01931"
196158..199526
/note="assembly_fragment:03255"
BASE COUNT 57428 a 38898 c 39349 g 63347 t 504 others
ORIGIN
```

alignment_scores:

```
Quality: 1607.00 Length: 301
Ratio: 5.339 Gaps: 0
Percent Similarity: 100.000 Percent identity: 100.000
```

alignment_block:

US-09-775-181-4 x AL139821

Align seg 1/1 to: AL139821 from: 1 to: 199526

```
1 MetGlyAlaMetAlaTyrProLeuLeuLeuCysLeuLeuLeuAlaGlnle 17
|||||
97158 ATGGAGGCCATGGCTTACCCCTCTACTCTCTGCTCTGCTTCAGCT 97207
|||||
17 ucGlyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAsps 34
|||||
97208 GGGATTGGAGCTGTTGGCCAGCGCGCACCCCAAGGAGCGCGGATT 97257
|||||
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50
|||||
97258 CCCCTCGAGAGAGAGACCCCGAAGGGGAAGCGCGCAGCCAGCCGGT 97307
|||||
```

```
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG1 67
|||||
97308 CGAGCCCTCTCGCTCGGACTCTCGCTCCCTGGAGCGCTCCACCGATGG 97357
67 yThrIleLeuAlaGlnLysLeuAlaGluGluValProMetAspValAlaS 84
|||||
97358 CACCATCTTGGCGCAGAACTCGCCGAGGAGTGCCTCATGACGTGGCCT 97407
84 eTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCysSer 100
|||||
97408 CTTACCTCTACACGGGGACTCCACAGCTGAGCGAGGCACTGCTCC 97457
101 GlyArgTyrGluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlaSe 117
|||||
97458 GGCCGCTTACAGTTGGCGGCTTGGCGGAGTGCACAGCCCTGGCCAG 97507
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
|||||
97508 CGCGCACCCCTCTCTGACCGGCGCTGGACACACTGACACAGCCACCA 97557
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
|||||
97558 ACTCTCTACAGTGTGCTGCAGAGCAATAAGTCGGGGAGCAGAACTTG 97607
151 GlnAspAspLeuAspTrpTyrGlnAlaLeuValTrpSerLeuLeuGluG1 167
|||||
97608 CAGGACGACCTGATGTTGGTACCAGCGCTGCTGTGAGCGCTTCTGGAGG 97657
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeuS 184
|||||
97658 CGAGCCGACGATCTCCGGCGGCGCATACCTTCAGCAGCGATTCGCTG 97707
184 eAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
|||||
97708 CCGCAGCGGCCCCACAGTCTTCTCTCAGGCCACGCGCGGAGGAGCCGC 97757
201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
|||||
97758 ATCTCTGCTCCAGACCTGTCTCTCCGACCCACCTGCGCAACGCCAC 97807
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
|||||
97808 TCTGGAGACGAGTGTGTCCACGGCTCCGGCGCAAGTGGAGGCGCCACT 97857
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
|||||
97858 TACACCGCGCGGCCCATCATCAGGGCGCCGGGCGCTGGCCACAGCTGG 97907
251 ArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSerPr 267
|||||
97908 CGCGCAAGCAGCGGCTCGCGGGGACAGAGCCACTTCAAGTGGTCTCC 97957
267 oProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuValT 284
|||||
97958 GCCTTATCTGGAGTGGAGAACGGGATTCACAGCCCGGGTGGCTGGTTA 98007
284 hrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluPhe 300
|||||
98008 CTCTTCTCTGCTGCTACGGGTTCAGCCTAACCTGTGCTCCGGATTC 98057
301 Arg 301
|||
98058 AGG 98060
```

seq_name: gb_hgt20:AL355587

seq_documentation_block:

LOCUS AL355587 151218 bp DNA HTG 15-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-561H23, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AL355587
VERSION AL355587.8 GI:13660940

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 151218)
Burton, J.
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Apr 17, 2001 this sequence version replaced gi:13624988.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA561H23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: pGAP4; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147891 bases at least Q40
Consensus quality: 149178 bases at least Q30
Consensus quality: 149893 bases at least Q20
Insert size: 150418; sum-of-contigs
Insert size: 151193; 7.4% error; agarose-fp
Quality coverage: 4.69x in Q20 bases; sum-of-contigs Quality
coverage: 5.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 21608: contig of 21608 bp in length
* 21709 21708: gap of 100 bp
* 21709 48919: contig of 27211 bp in length
* 48920 49019: gap of 100 bp
* 49020 57539: contig of 8520 bp in length
* 57540 57639: gap of 100 bp
* 57640 69961: contig of 12322 bp in length
* 69962 70061: gap of 100 bp
* 70062 101299: contig of 31238 bp in length
* 101300 101399: gap of 100 bp
* 101400 114233: contig of 12834 bp in length
* 114234 114333: gap of 100 bp
* 114334 126584: contig of 12251 bp in length
* 126585 126684: gap of 100 bp
* 126685 138216: contig of 11532 bp in length
* 138217 138316: gap of 100 bp
* 138317 151218: contig of 12902 bp in length.
Location/Qualifiers
1. 151218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-561H23"
/clone_lib="RPC1-11.2"
1. 21608
/note="assembly_fragment:00445
fragment_chain:1
clone_end:T7
vector_side:left
21709. 48919
/note="assembly_fragment:00191
fragment_chain:1
49020. 57539
/note="assembly_fragment:01755


```
misc_feature      fragment_chain:1"
57640..69961
/note="assembly_fragment:00689
fragment_chain:1"
70062..101299
/note="assembly_fragment:01153
fragment_chain:2"
101400..114233
/note="assembly_fragment:01618
fragment_chain:2"
114334..126584
/note="assembly_fragment:02159
fragment_chain:2"
126685..138216
/note="assembly_fragment:02086
fragment_chain:2"
138317..151218
/note="assembly_fragment:00492
fragment_chain:2
clone_end:SP6
vector_side:right"
BASE COUNT 45266 a 28197 c 29184 g 47771 t 800 others
ORIGIN

alignment_scores:
  Quality: 489.00      Length: 253
  Ratio: 3.075         Gaps: 11
  Percent Similarity: 62.846 Percent Identity: 50.198

alignment_block:
US-09-775-181-4 x AL355587 ..
Align seg 1/1 to: AL355587 from: 1 to: 151218
352 CysIleCysLysAlaGlyPheThrHisProGlyValLeuPro.ValAsnA 368
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66126 TCATTAAGCCAGCAGCAGCGACGTCGCTGGAGTCTCAGCTACTTGGAA 66175
368 snPheArgArgGlyProAspGlnHisIleSerGlySerThrLysAsp 384
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66176 GGCTGAGGCGA.....GAGGAATGCTTGAGCCAGAGTTCAGGAT 66216
385 ValSerGluGluAlaTyrValCysLeu...ProCysArgGluGlyCysPr 400
||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66217 GCAGTACACTATGATGCTGTGTAATAGCCACTGCCTCCAGCATGGG 66266
400 oPheCysAlaAspAspSerProCysPheValGlnGluAspLysTyr...L 416
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66267 CAACATAGCG.....AGACCTGTCTCTACACAAACAAACAAATACAAA 66310
416 euArgLeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeuAspPhe 432
||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66311 CAAAGCAAACTCTGATTACTGG..... 66333
433 ValSerMetLeuValValTyrHisPheArg.LysAlaLysSerIleArgA 449
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66334 .....ATTTGGTGGTTTACATTTTACCTAGAAAA..... 66364
449 laSerGlyLeuLeuGluThrIleLeuPheGlySerLeuLeuLeu 465
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66365 .....ATAGTTATTTAGAG.....TTA 66382
466 TyrPheProValIleLeuTyrPheGluProSerThrPheArgCysIle 482
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66383 TACTTCTCAATGTTAGGTTGCTTTC..... 66409
482 cLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrGlyThrV 499
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66410 .....ACAT 66413
499 aThrLeuLysLeu.HisArgValLeuLysValPheLeuSerArgThrAl 515
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
```

```
56414 TTGAACCTTTTGTCTTTTCTAGGCTTTTGAAGGTGTTTCTTTCACGAACGCC 66463
515 acGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLeuAlaVal 532
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66464 TCACACGAATTCATATATGACTGGCGGACGGTTCATGAGGATGCTGGCAG 66513
532 alIleLeuLeuValValPheTrpPheLeuIleGlyTyrTrpThrSerSerVal 548
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66514 TAATACTCTTGTGTAGTGTGTTTGTGTTTCTCATTTGGCTGGACTTCATCTGTG 66563
549 CysGlnAsnLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSe 565
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66564 TGCCAGTATTTCGACAAACAGATTTTCACATTATGGCCATGGGAAACATC 66613
565 rAspHisLeuIlePheAsnMetCysLeuIleAspArgTrpAspTyrMetT 582
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66614 CGATCACCTCATCTTCAATATATGCTCATTTGACCGCTGGGACTACATGA 66663
582 hrAlaValGlyMetTrpSerLeuValSerTyrAspGlyLeuThrIlePhe 598
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
66664 CAGCAGTTGGTATGTGTCACCTGTTTGTGATGATGATGGTCTTACCATTTT 66713
599 Gln 599
|||||
66714 CAG 66716
seq_name: gb_htg21:AL359974
seq_documentation_block:
LOCUS AL359974 143878 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 10 clone RP13-236A4, *** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION AL359974
VERSION AL359974.4 GI:9864577
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143878)
AUTHORS Sims S.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:9795088.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BB236A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 130224 bases at least Q40
Consensus quality: 135875 bases at least Q30
Consensus quality: 138890 bases at least Q20
Insert size: 141478; sum-of-contigs
Quality coverage: 205930; 23.5% error; agarose-fp
Quality coverage: 3.17x in Q20 bases; sum-of-contigs Quality
coverage: 2.50x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 12130: contig of 12130 bp in length
* 12131 12230: gap of 100 bp
* 12231 20274: contig of 8044 bp in length
* 20275 20374: gap of 100 bp
* 20375 23062: contig of 2688 bp in length
* 23063 23162: gap of 100 bp
* 23163 26227: contig of 3065 bp in length
* 26228 26327: gap of 100 bp
* 26328 33260: contig of 6933 bp in length
* 33261 33360: gap of 100 bp
* 33361 35571: contig of 2211 bp in length
* 35572 35671: gap of 100 bp
* 35672 38934: contig of 3163 bp in length
* 38935 44371: contig of 5437 bp in length
* 44372 44471: gap of 100 bp
* 44472 47729: contig of 3258 bp in length
* 47730 47829: gap of 100 bp
* 47830 52657: contig of 4828 bp in length
* 52658 52757: gap of 100 bp
* 52758 54776: contig of 2019 bp in length
* 54777 54876: gap of 100 bp
* 54877 57327: contig of 2451 bp in length
* 57328 57427: gap of 100 bp
* 57428 60599: contig of 3172 bp in length
* 60600 60699: gap of 100 bp
* 60700 64861: contig of 4162 bp in length
* 64862 64961: gap of 100 bp
* 64962 71930: contig of 6969 bp in length
* 71931 72030: gap of 100 bp
* 72031 75429: contig of 3399 bp in length
* 75430 75529: gap of 100 bp
* 75530 78541: contig of 3012 bp in length
* 78542 78641: gap of 100 bp
* 78642 80961: contig of 2320 bp in length
* 80962 81061: gap of 100 bp
* 81062 83099: contig of 2038 bp in length
* 83100 83199: gap of 100 bp
* 83200 95484: contig of 12285 bp in length
* 95485 95584: gap of 100 bp
* 95585 97735: contig of 2151 bp in length
* 97736 97835: gap of 100 bp
* 97836 102251: contig of 4416 bp in length
* 102252 102351: gap of 100 bp
* 102352 131392: contig of 29041 bp in length
* 131393 131492: gap of 100 bp
* 131493 137581: contig of 6089 bp in length
* 137582 137681: gap of 100 bp
* 137682 143878: contig of 6197 bp in length.
TUES
source
1. 143878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPCI-13.1"
1. 12130
/notes="assembly_fragment:01047
fragment_chain:1"
12231..20274
/notes="assembly_fragment:00007
fragment_chain:1"
20375..23062
/notes="assembly_fragment:01237
fragment_chain:1"
23163..26227
/notes="assembly_fragment:01103
fragment_chain:1"
26328..33260
/notes="assembly_fragment:00266
fragment_chain:1"
33361..35571
/notes="assembly_fragment:01006
fragment_chain:1"
misc_feature
/notes="assembly_fragment:00201
fragment_chain:2"
misc_feature
/notes="assembly_fragment:01057
fragment_chain:2"
misc_feature
/notes="assembly_fragment:00179
fragment_chain:2"
misc_feature
/notes="assembly_fragment:00444
fragment_chain:3"
misc_feature
/notes="assembly_fragment:00414
fragment_chain:3"
misc_feature
/notes="assembly_fragment:00586
fragment_chain:4"
misc_feature
/notes="assembly_fragment:00641
fragment_chain:4"
misc_feature
/notes="assembly_fragment:00717
fragment_chain:5"
misc_feature
/notes="assembly_fragment:01116
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00156"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00233"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00273"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00279"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00515"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00654"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00675"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00719"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00767"
fragment_chain:5"
misc_feature
/notes="assembly_fragment:00852
fragment_chain:5"
BASE COUNT 44271 a 25988 c 25788 g 45414 t 2417 others
ORIGIN
alignment_scores:
Quality: 409.00 Length: 108
Ratio: 4.544 Gaps: 2
Percent Similarity: 83.333 Percent Identity: 73.148
alignment_block:
US-09-775-181-4 x AL359974/rev ..
Align seg 1/1 to reverse of: AL359974 from: 1 to: 143878
372 ArgGlyProAspGlnHisIleSerGlySerThrLysAspValSerGluG1 388
|||||
89283 AGGGGCCGGATCAGCATATTTCCAGGAAGTACAAAGATGTGTCAAGA 89234
388 uAlaTyrValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
|||||
89233 AGCCTATGCTGCTACCTTGACGGAGGCGCTGCCCTTCTGTGCTGATG 89184
405 sPserProCysPheValGlnGluAspLysTyrLeuArgLeuAlaIleIle 421
```

```

89183 ACAGCCATGCTCGTCCAGGAGATAGTATTTAGCACTTGCATCATC 89134
|||||
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
89133 TCCTTCAAGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 89084
|||||
438 lTyrlHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuLeuLeu 455
|||||
89083 CTACCACTTTCGCAAGCAAG...GTAACCCAGGAGCAACCTGGTT... 89041
|||||
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyrlPheProValValIle 471
|||||
89040 .....ATGATCTGTATTACAGACCACTCT 89014
|||||
472 LeuTyrlPheGluProSerThrPhe 479
|||||
89013 TATCTTTTAAAGCAACAGAAATTT 88990
|||||
seq_name: gb_hgt20:AL161654

seq_documentation_block:
LOCUS AL161654 159596 bp DNA HTG 11-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-59G22, *** SEQUENCING IN
PROGRESS ***, 8 unordered pieces.
ACCESSION AL161654
VERSION AL161654.8 GI:13620309
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton, J.
Direct Submission
Submitted (10-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:13567947.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA59G22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 157954 bases at least Q40
Consensus quality: 158428 bases at least Q30
Consensus quality: 158729 bases at least Q20
Insert size: 158896; sum-of-contigs
Insert size: 159333; 6.0% error; agarose-fp
Quality coverage: 6.69x in Q20 bases; sum-of-contigs Quality
coverage: 6.67x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 24276: contig of 24276 bp in length
* 24277 24376: gap of 100 bp
* 31732 31732: contig of 7356 bp in length
* 31733 31832: gap of 100 bp
* 31833 67342: contig of 3510 bp in length
* 67343 67442: gap of 100 bp
* 70588: contig of 3146 bp in length

```

```

* 70589 70688: gap of 100 bp
* 70689 81295: contig of 10607 bp in length
* 81296 81396: gap of 100 bp
* 81396 111898: contig of 30503 bp in length
* 111899 111998: gap of 100 bp
* 111999 126181: contig of 14183 bp in length
* 126182 126281: gap of 100 bp
* 126282 159596: contig of 33315 bp in length.
FEATURES
            Location/Qualifiers
                1..159596
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="RP11-59G22"
                /clone_lib="RPC1-11.1"
                1..24276
                /note="assembly_fragment:01962"
                fragment_chain:1
                clone_end:17
                vector_side:left
                24377..31732
                /note="assembly_fragment:01615"
                fragment_chain:1
                31833..67342
                /note="assembly_fragment:02112"
                fragment_chain:1
                67443..70588
                /note="assembly_fragment:00781"
                fragment_chain:1
                70689..81295
                /note="assembly_fragment:00892"
                fragment_chain:1
                81396..111898
                /note="assembly_fragment:02522"
                fragment_chain:1
                111999..126181
                /note="assembly_fragment:01840"
                fragment_chain:1
                126282..159596
                /note="assembly_fragment:01914"
                fragment_chain:1
                clone_end:SP6
                vector_side:right
BASE COUNT 54172 a 29996 c 28660 g 46068 t 700 others
ORIGIN
alignment_scores:
    Quality: 403.00      Length: 108
    Ratio: 4.478        Gaps: 2
    Percent Similarity: 83.333 Percent Identity: 72.222
alignment_block:
US-09-775-181-4 x AL161654/rev ..
Align seg 1/1 to reverse of: AL161654 from: 1 to: 159596
372 ArqGlyProAspGlnHisIleSerGlySerThrLysAspValSerGluG1 388
|||||
102605 AGGGTCCGATCAGCATATTTTCAGGAAGTACAAAGATGTGTCAGAAGA 102556
388 uAlatyrValCysLeuProCysArgGluGlyCysProPheCysAlaAspA 405
|||||
102555 AGCCTATGCTGCCTACCTTGCAGGAGGCGTCCCTCTCTGTGCTGATG 102506
405 sPSerProCysPheValGlnGluAspLysIleuArgLeuAlaIleIle 421
|||||
102505 ACAGCCATGCTCGTCCAGGAGATAGTATTTAGCACTTGCATCATC 1024456
422 SerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValVa 438
|||||
102455 TCCTTCAAGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 102406

```

```
438 lTyHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuL 455
|||||
102405 CTACCACTTTGCAGCAAG...GTAAACCCAGACCTGGTT... 102363
455 euGluThrIleLeuPheGlySerLeuLeuLeuTyPheProValValIle 471
|||||
102362 .....ATGATCCGCTGATTACAGAGCAACCTCT 102336
472 LeuTyPheGluProSerThrPhe 479
|||||
102335 TATCTTTTAAACCAACAGATTT 102312
seq_name: gb_htg5:AC017271

seq_documentation_block:
LOCUS AC017271 10892 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017271
VERSION AC017271.1 GI:6553715
FEATURES
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 10892)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210111 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
1. 10892
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3383 a 2213 c 2125 g 3171 t
ORIGIN

alignment_scores:
Quality: 259.50 Length: 414
Ratio: 1.359 Gaps: 21
Percent Similarity: 46.135 Percent Identity: 25.845

alignment_block:
US-09-775-181-4 x AC017271/rev ..
Align seg 1/1 to reverse of: AC017271 from: 1 to: 10892
226 LeuArgLysTrpArg...Pro.HisLeuHisArgArgGlyProAsnG 241
|||||
9487 CTTTCAGAGTTGGCGCTTCATTGCCCGCAGGAGGATGTGTC 9438
241 lnGlyProArgGlyLeuGlyHisSerTrpArgLysAspGlyLeuGly 257
|||||
9437 AA.....CGATGGCTCGAGGAGTTTCGG..... 9412
258 GlyAspLysSerHisPheLysTrpSerProTyPheLeuGluCysGluAs 274
|||||
9411 .....TCGCCCTCATGGGTGAGTTGGCCA 9386
274 nGlySerTyLysProGly.....TrpLeuValThrLeuSers 287
|||||
9385 AACTTTATACGATAATTTCGTGACATTAGTTCTGGATAATA..... 9346
287 erAlaIleTyGlyLeuGlnProAsnLeuValProGluPheArgGlyVal 303
|||||
9346 .....TATATAATACCATATCTCTC 9328
304 MetLysValAspIleAsnLeuGlnLysValAspIleAsnSerGlnCysSerSe 320
|||||
9327 .....TGTAGTTG 9320
320 rAspGlyTrpPheSerGlyThrHisLysCysHisLeuAsnAsnSerGlnC 337
|||||
9319 CGAT.....CGGAATACGACCTTCTGCTCCTCACCAGAACAAAC 9279
337 yMetProIleLysGlyLeuGlyPheValLeuGlyAlaTyGluCysIle 353
|||||
9278 CCGCCGCCACCCGGGATGTG.....TACACCTGCTG 9247
354 CysLysAlaGlyPheTyHisProGlyValLeuProValAsnAsnPheAr 370
|||||
9246 TCGCGGAATCCTACTACTCCCAAT...TCCAGCTCCAGGATTCGG 9200
370 gArgGlyProAspGlnHisLysSerGlySerThrLysAspValSerG 387
|||||
9199 T.....GGCATCGAGTGAACCTGCTCG 9177
387 lu.....GluAlaTyValCysLeuProCysArgGluGlyCysProphe 401
|||||
9176 AGGCTACGACAACTACTCTGCTCCGTCCTCCCGCGGATGCACCTAAC 9127
402 CysAlaAspAspSerProCys.....PheValGlnGlu..... 412
|||||
9126 TCGCATAGCAACGGCTCTCTGTCACCTTCAGGAGGAGGAGGTGCTCAA 9077
413 ....AspLysTyLeuArgLeuAlaIleSerPheGlnGlyLeuCysM 428
|||||
9076 TGTGGACGCTGCTGCGCTCTCTGCTGCGCATCGTACTGGCGCCTGCA 9027
428 etLeuLeuAspPheValSerMetLeuValValTyHisPheArgLysAla 444
|||||
9026 TCCTGTCTGCATGTCTCCTCGGGTGATTCTTCCGGCAGCGAAAGTGC 8977
445 LysSerIleArg..... 448
8976 AAGTATTTAGTTTACATATCCTAAAGTTGGAAACTTTAGTAATCCTT 8927
449 .....AlaSerGlyLeu...IleLeuLeuGluThrI 458
8926 TCATCATTTTCAGGCCATTGCGTCTGTGTATGTGGACTGTCTGGAGACGA 8877
458 leLeuPheGlySerLeuLeuTyPheProVal.ValIleLeu..... 472
8876 TACTGTGGGCATGTTTACTTTATGCATCTGTGAGTGTCTCTGGCGAT 8827
472 ..... 472
8826 GTTCACCTTCAACTTCTGTGTATTATATATCATCTTTAGTTGCCG 8777
473 .....TyPheGluProSerThrPheArgCysIleLeuLeuArgTrpAla 487
8776 TCATTTCTTCCCGATCCACCGAGCGCTCTCTGGAGCCTGGCTT 8727
488 ArgLeuLeuGlyPheAlaThrValTyGlyThrValThrLeuLysLeuHi 504
8726 CCGGAGCTGGGCTTCATCCTCTGCTACGGCGCCATCATCATGAAGTTGTA 8677
504 sArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyM 521
8676 CCGCCACCTGTGGACTTTCGCACCCCGAAAGCGCATCGCTGG...GTGC 8630
521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuValVal 537
8629 TCGCGCAGCTGGACCTGCTCAAGTATCTGGCCACCATGCTTTCGCGTC 8580
538 PheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGluLy 554
8579 ATCTGTACATGCGCCCTTTACGGCTCTGCTGTGATCTTCTCGAAAG 8530
```



```
10865 AAGTATTTAGTTTACATATCCCTAAAGTTGGGAACCTTTGAGTAATCCCTT 10816
449 .....AlaSerGlyLeu...IleLeuLeuGlnThrI 438
10815 TCATCATTTTTCAGGCCATTCGCTGTGTATGTGAGCTGCTGCTGAGACGA 10766
458 leLeuPheGlySerLeuLeuTyPheProVal.ValIleLeu..... 472
10765 TACTGTGGGCATGTTTATCTTATGCACTCTGTGAGTCTGCTGCGCAT 10716
472 ..... 472
10715 GTCACCTTCAACTCTGTGCTGTATTACATATTCATCTTAGTTGCCG 10666
473 ....TyrPheGluProSerThrPheArgCysIleLeuLeuArgTTrpAla 487
10665 TCCATTCTTTCGCCGATCCACCGAGCGTGCCTTCTGGAGCCCTGGCTT 10616
488 ArgLeuLeuGlyPheAlaThrValTyGlyThrValThrLeuLeuIleH 504
10615 CGGGAGCTGGGCTTCATCAGCTGCTACGGCGCATCATCTGAAGTTGTA 10566
504 sArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyCM 521
10565 CCCCACCTGTGGACTTTCGCCACCCGGAAGCGCATCGCTGG...GTGC 10519
521 etThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuVal 537
10518 TGGCGACGTGGACCTGCTCAAGTATCTGGGCACCATGCTTTCGCGTC 10469
538 PheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGlu 554
10468 ATCTGTATCATGCGCCCTTTACGGCTCGTGTGATCTTCTCGAAG 10419
554 s...GlnIleSerLeuIleGlyGlnGlyThrSerAspHisLeuIleP 570
10418 TGGCGACGTGGAGAGCTCAGGGAGCGGACAGC..... 10385
570 heAsnMetCysLeuIleAspArgTrpAspTyMetThr 582
10384 ..AACACCTGCCATCGCTCAAGTGGGAAGTGTGCACG 10349
seq_name: gb_inl:AE003575
seq_documentation_block:
LOCUS AE003575 327209 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386046 section 16
of 16, complete sequence.
VERSION AE003575 AE002638
WORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 327209)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.C., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Wan,K.H., Doyle,C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Fannkotch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,M., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferrera,C., Ferrieras,S.,
Fleischmann,W., Posler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moberg,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nuskern,D.R., Pacleb,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 327209)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7295650.
COMMENT
FEATURES
source
1. 327209
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/gene="CG11933"
/product="CT35796"
/db_xref="FLYBASE:FBan0011933"
/db_xref="FLYBASE:FBgn0031615"
/evidence=not_experimental
<29815..>30955
/gene="CG11933"
/map="25A3-25A3"
/db_xref="FLYBASE:FBan0011933"
/db_xref="FLYBASE:FBgn0031615"
/evidence=not_experimental
join(29815..29907,30773..30955)
/gene="CG11933"
/note="CG11933 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0011933"
/db_xref="FLYBASE:FBgn0031615"
/evidence=not_experimental
/protein_id="AAF50997.1"
/db_xref="GI:7295691"
/translation="MAQPLFGSPDCHLWTGLGSGCGFCGSSGCCDYLRLHLSFNFQD
YWQARVAREYQIFPGQMPLOYGVHDLRLFLPTNCQEYNSNQAVQ"
complement(<31022..>31300)
/gene="CG11932"
/product="CT35794"
/db_xref="FLYBASE:FBan0011932"
/db_xref="FLYBASE:FBgn0031616"
/evidence=not_experimental
complement(<31022..>31300)
/gene="CG11932"
/map="25A3-25A3"
/db_xref="FLYBASE:FBan0011932"
/db_xref="FLYBASE:FBgn0031616"
/evidence=not_experimental
complement(31022..31300)
/gene="CG11932"
```



```
370 gArgArgGlyProAspGlnHisIleSerGlySerThrLysAspValSerG 387
260158 T.....GCGATCGAGTGAACACTGCCG 260180
387 lu.....GluAlaTyrValCysLeuProCysArgGluGlyCysProPhe 401
260181 AGGCTACGACAACTACTCGTGCATTCGCTGCCGCGGATGCACATAAC 260230
402 CysAlaAspSerProCys.....PheValGlnGlu..... 412
260231 TGGCATAGCAACGGCGTCCTGTCGACCTTTCAGGAGGAGGAGGTGCTCAA 260280
413 ...AspLysTyrLeuArgLeuAlaIleIleSerPheGlnGlyLeuCysM 428
260281 TGTGGAGCGCTGTCTGCCCTCTCGTGGCCATCGTACTGGCGCGCTGCA 260330
428 etLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysAla 444
260331 TCCTGTGCTGCATGTCTCGGGGTGATTGCTTCGGGACGCGAAAGTGC 260380
445 LysSerIleArg..... 448
260381 AAGTATTAGTTTACATATCTTAAGTTGGGAACTTTGAGTAAATCCTT 260430
449 .....AlaSerGlyLeu.....IleLeuLeuGluThrI 458
260431 TCATCATTTTTCAGGCCATTCGCTGCTGATGTGCTGCTGCTGACGACGA 260480
458 leLeuPheGlySerLeuLeuTyrPheProValValIleLeu..... 472
260481 TACTGTGGGCATGTATTACTTATGCACTCTGTGAGTCTCTCGCGGAT 260530
472 ..... 472
260531 GGTCACTTCAACTCTGTGTGATTTACATATTCATCTTTAGTTGCCG 260580
473 ....TyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAla 487
260581 TCCATTCTTTCCCGCATTCACCGAGCGCTGCTTCTGGAGCGCTGGCTT 260630
488 ArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHi 504
260631 CGGAGCTGGCTTTCATCATCTCTGAGCGGCATCATCTGAAGTTGTA 260680
504 sArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyrM 521
260681 CGCCACCTGGTGGACTTTCGACCGCGGAAAGCGCATCGCTGG...GTGC 260727
521 etThrGlyArgValMetArgMetLeuAlaValIleLeuLeuValVal 537
260728 TGGCGGACCTGGACCTGCTCAAGTATCTGGGACCATGCTTCTTGGCCGTC 260777
538 PheTrpPheLeuIleGlyTrpThrSerValCysGlnAsnLeuGluLys 554
260778 ATCTGTACATGCGCGCTTACGGCTCGTGTGATCTCTTCGAAAG 260827
554 s...GlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisLeuIleP 570
260828 TGGCGACCTGGAGACCTTCAGGGAGCGGACACG..... 260861
570 heAsnMetCysLeuIleAspArgTrpAspTyrMetThr 582
260862 ..AACACCTGCCATCGCTCAAGTGGGAACTGGTCACG 260897
```

seq_name: gb_htg20:AL354976

seq_documentation_block:

LOCUS AL354976 186964 bp DNA 07-APR-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-2016, *** SEQUENCING IN
PROGRESS ***, 4 unordered pieces.
ACCESSION AL354976
VERSION AL354976.7 GI:13561182
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186964)
Plumb.B.
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 8, 2001 this sequence version replaced gi:12330777.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba2016
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 185320 bases at least Q40
Consensus quality: 186184 bases at least Q30
Consensus quality: 186460 bases at least Q20
Insert size: 186664; sum-of-contigs
Quality coverage: 7.22x in Q20 bases; agarose-fp
coverage: 8.22x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 149917: contig of 149917 bp in length
* 149918 150017: gap of 100 bp
* 150018 153565: contig of 3548 bp in length
* 153566 153685: gap of 100 bp
* 153686 162836: contig of 9171 bp in length
* 162837 162936: gap of 100 bp
* 162937 186964: contig of 24028 bp in length.

FEATURES
..source

Location/Qualifiers
1..186964
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-2016"
/clone_lib="RPC1-11.1"

misc_feature
1..149917
/note="assembly_fragment:00886.0"

misc_feature
150018..153565
/note="assembly_fragment:01186"

misc_feature
153666..162836
/note="assembly_fragment:01998"

misc_feature
162937..186964
/note="assembly_fragment:00147"

clone_end:SP6

vector_side:right

BASE COUNT 64194 a 34055 c 33553 g 54858 t 304 others
ORIGIN

alignment_scores:

Quality: 202.00 Length: 42
Ratio: 5.179 Gaps: 0

Percent Similarity: 92.857 Percent Identity: 85.714

alignment_block:

US-09-775-181-4 x AL354976

56355TACCACATTCGCCAA 56341

seq_name: gb_inl:AC007772

seq_documentation_block:

LOCUS AC007772 186607 bp DNA INV 24-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 3R, region 92E-93A, BAC clone
BACR19p12, complete sequence.

ACCESSION AC0099347

VERSION AC007772.5 GI:13122713

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 186607)

AUTHORS

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Farrera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattoti, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Shapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of drosophila chromosome 3R, region 92E-93A

Unpublished

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 186607)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champs, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomoltan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

Direct Submission

Submitted (10-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 24, 2001 this sequence version replaced gi:6957932

gi:6598746.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bugbefruitfly@berkeley.edu.

FEATURES

source

1. 186607
/organism="Drosophila melanogaster"
/strain="v: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="92E-93A"
/clone="BACR19p12 (D739)"
/clone.lib="RpCl-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACs:6)"

BASE COUNT 53882 a 40666 c 38315 g 53744 t

ORIGIN

alignment_scores:

Quality: 183.50 Length: 511
Ratio: 0.913 Gaps: 28
Percent Similarity: 39.335 Percent Identity: 22.701

alignment_block:
US-09-775-181-4 x AC007772/rev ..
Align seg 1/1 to reverse of: AC007772 from: 1 to: 186607

7 ProLeuLeuLeuCysLeuLeu.....LeuAlaGlnLeuGlyLeuGlyAl 21
46167 CCCTGCTCATCTCGAGATATGGCCCTTAGCAGAGCTTCTTCTGAGC 46118
21 aValGlyAlaSerArg.....AspProGlnGlyArgProAspS 34
46117 TACATCCTCCAGTCGAGGTTTATTGACGCCGCAACACGACACCTACGA 46068
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnClnProGly 50
46067 TCCCGCGATGATGATCTACTTTCTCTCCACCGTACCGATGATCCG... 46021
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAsp.G 67
46020CCAAACCCACATATCAACGCCCTCG 45998
67 lyThrIleLeuAlaGlnLysLeuAlaGluGluValPrometaspValAla 83
45997 GCCGTGTACTT.....CTCCCCCAACAGCTGCTATTGCTGCTGCTA 45957
84 SerTyLeuTyThrGlyAspSerHisGlnLeuLysArgAlaAsn..... 98
45956 TCGCGGCTTCTTCAATAAGACGTTCCCGAGATTTCGGGCCAAGACCTTCA 45907
98ATCCAAAGG 45802
45906 GGCTGGACGACTTCAACGATCCCATTCATCTCCCAAGATATCGACGTGG 45857
99CysSerGlyArgTyGluLeuAlaGlyLeuProGlyLys 111
45856 AATACTTTTCGATGTTTCAGGA.....TCTGGGCGCCATCA... 45822
112 TrpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspTh 128
45821ATCCAAAGG 45802
128 rLeu.ThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLys 144
45801 ACTACACCCACGACCTGTATAAAATAAAC..... 45773
145 SerArgGluGlnAsnLeuGlnAspLeuAspTrpTyTrpGlnAlaLeuVa 161
45772GAGTGGTACCGCGCTGGCT 45753
161 lTrpSerLeuLeuGluGluProSerIleSerArgAlaAlaIleThrP 178
45752 ACCAGAACACGTCGAGGA..... 45734
178 heSerThrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAla 194
45734 45734
195 ThrArgGluGluSerArgIleLeuLeuGlnAspLeuSerSerAlaPr 211
45733 ...CGGCACGATACGAGATCACCTACCAG.....GTGGAAT 45699
211 oHisLeuAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuArga 228
45698 CGGCTATCGCAACAACAAACAGACGATATACCTTCCAC..... 45659
228 rgLysTrpArgProHisLeuHisArgArgGlyPro..... 239
45658GGACCCCTGGCTCTCGAAGAA 45638

```
240 AsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLe 256
||| |||||
45637 AACCTGCTCGC..... 45626

256 uGlyGlyAspLysSerHisPheLysTrpSerProTyrLeuGluCysG 273
|||||:|||||
45625 .....ATTAAATTACAGGCCGCTACTTGGATGT. 45596

273 LuAsnGlySerTyrLys.ProGlyTrp..... 281
|||:|||||
45595 .....GCAGGTCCAACAGTGGCTGGTGGCCGAGTAGTCCAAATGGC 45551

282 .....LeuValThrLeuSer.....SerA 288
|||||:|||||
45550 GATATCTACCCCGACACAGCGAGTTCCTCCGTCACATTGAGTATCCCAAGTA 45501

288 LalleTyrGlyLeuGlnProAsnLeuValPro..... 298
|||:|||||:|||||
45500 AGATACCTTGAATATCCCGTGAATACCTCTTTTATCTACTGATCGCT 45451

299 ...GluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
|||||:|||||
45450 TTTAGATACACGCCGCTTCCGTTCTTGAGATGAGCTTCGAGCGTATCGA 45401

314 PileAspGlnCys.....SerSerAspGlyTrpPheS 325
|||||:|||||
45400 CATAAACCAAGTGCATTCGCTGAAGCAACAAGGACCTTAATCACCTTG 45351

325 erGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProIleLys 341
|||:|||||:|||||
45350 CGGATACCGCGCGGTGTAAAGAAACGACGAGTGTGAACCATATACAA 45301

342 GlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCysLysAlaGlyPh 358
|||||:|||||
45300 GCCTGGCGCTTAGCGCGGTGGCTACCGCTGCTGTAAGCCAGCTTT 45251

358 eTyrHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
|||:|||||
45250 TCGCTGCCCAAGCTAGTG.....CGCGCACCTTATCTGG 45216

375 spGlnHisIleSerGlySerThrLysAspValSerGluGluAlaTyrVal 391
|||:|||||:|||||
45215 GCAGATGTGGAGCGCGGCTCGGCAACAGTACTACACGAGTAGCAC 45166

392 CysLeuProCysArgGluGlyCysProPhe.....CysAlaAs 404
|||||:|||||
45165 TGCCTT.....AAGATGCGCTCTATCTTTAAGTAGCAATATGTAAG 45122

404 p.....AspSerProCysPheValGlnGluAspLysTyrLeuArgL 418
|||:|||||
45121 TATGAGATTTTCACTCTTGATGTTTT..... 45097

418 euAlaIleSerPheGlnGlyCysMetLeuLeuAspPheValSer 434
|||||:|||||
45096 ....TTTTTAGGGATCCAAAGCTTCCCATTCAGTGGGATAAGGCTCC 45052

435 MetLeuValTyrHisPheArgLys 443
|||||:|||||
45051 .....TACCACATTCGCCAA 45037
```

seq_name: gb_inl:AC007771

```
seq_documentation_block:
LOCUS AC007771 197071 bp DNA INV 15-FEB-2001
DEFINITION Drosophila melanogaster, Chromosome 3R, region 92E-92F, BAC clone
BACR26M03, complete sequence.
ACCESSION AC007771.7 GI:12831351
VERSION AC007771.7
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
```

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS

1 (bases 1 to 197071)
Celisner, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fieriera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Shanley, M., Strong, R., Svirkas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 92E-92F

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 197071)
Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Madad, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (10-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 15, 2001 this sequence version replaced gi:7018763.

Sequence submitted by:

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgpe@fruitfly.berkeley.edu.

FEATURES

source

1..197071
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="92E-92F"
/clone="BAC BACR26M03 (D738)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"

BASE COUNT 54696 a 44041 c 42886 g 55448 t

ORIGIN

alignment_scores:

Quality: 183.50 Length: 511
Ratio: 0.913 Gaps: 28
Percent Similarity: 39.335 Percent Identity: 22.701

alignment_block:

us-09-775-181-4 x AC007771/rev ..

Align seg 1/1 to reverse of: AC007771 from: 1 to: 197071

7 ProLeuLeuLeuCysLeuLeu.....LeuAlaGlnLeuGlyLeuGlyAl 21

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171917 CCTGCTGCATCTGCAGATATGGCCCTTAGCAGAGCTTCTTCTGGAGC 171868

21 aValGlyAlaSerArg.....AspProGlnGlyArgProAsps 34

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171867 TACATCTCCAGTCGAGGTTTATTCGACCCGCCATCAACGACACCTACGA 171818

34 erProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGly 50

```
171817 TCCCGCATGATGACTACTTCTCTCCACCGTAGCCGATGATCCG... 171771
171818 TCCCGCATGATGACTACTTCTCTCCACCGTAGCCGATGATCCG... 171771
51 ArgAlaSerAlaSerAspSerAlaProTrpSerArgSerThrAsp.G 67
171770 .....CCACCCACATATCAACGCCCTCG 171748
67 lyThrIleLeuAlaGlnLysLeuAlaGluGluValProMetAspValAla 83
171747 GCCGTGTAATT.....CTCCCCAACAGCTCGTATTCTCGTGTGCGTA 171707
84 SerTyrlLeuTyThrGlyAspSerHisGlnLeuLysArgAlaAsn..... 98
171706 TCGCGGCTTCTTCAATAAGAGCTTCCCGCAGATTCGGGCCAAGAACCCTTCA 171657
98 ..... 98
171656 GGCTGGAGCACTTCAACGATCCCATTCATCTGCAGAGATATCGACGTGG 171607
99 .....CysSerGlyArgTyrgluLeuAlaGlyLeuProGlyLys 111
171606 AATACTTTTCGATGTTCAGGA.....TCTGGCGGCCCATCA... 171572
112 TrpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspTh 128
171571 .....ATCCAAAG 171552
128 rLeu.ThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLys 144
171551 ACTACACCCAGCACTGTGTAATAATAAC..... 171523
145 SerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrgluAlaLeuVal 161
171522 .....GAGTGTACCGCGCTGGCT 171503
161 lTrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAlaIleThrP 178
171502 ACCAGACAACGTCGAGGA..... 171484
178 heSerThrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAla 194
171484 ..... 171484
195 ThrArgGluGluSerArgIleLeuLeuGlnAspLeuSerSerAlaPr 211
171483 ...CGGCAGCATAGAGATCACCTACCAG.....GTGGAAAT 171449
211 oHisLeuAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuArgA 228
171448 CCGCTATCGGAACAACAACACGAGAGATATACCTTCAC..... 171409
228 rgLysTrpArgProHisLeuHisArgArgGlyPro..... 239
171408 .....GACCGCTGGCTCTCAAGAA 171388
240 AsnGlnGlyProArgGlyLeuGlyHisSerTrpArgArgLysAspGlyLe 256
171387 AACCTGTGTCCG..... 171376
256 uGlyGlyAspLysSerHisPheLysTrpSerProProTyrlLeuGluCysG 273
171375 .....ATTAAATTACAAGGCCGTACTTCGATGTT. 171346
273 luAsnGlySerTyrlLys.ProGlyTrp..... 281
171345 .....GGCAGTCCAACAAGTGGCTGGTGGCGGCAGTAGTCCCAATTCGCG 171301
282 .....LeuValThrLeuSer.....Sera 288
171300 GATATCTACCCCGCACACGACGAGTTCGTCACATTGAGTATCCCAAGTA 171251
288 laIleTyrlGlyLeuGlnProAsnLeuValPro..... 298
```

```
171250 AGATACCTTGAATATCCCTCAATACCTCTCTTTTATCTACTGTATCGCT 171201
299 ...GluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
171200 TTATAGATACACGCCGCTTTCGTTCTTGAGATGGACTTCGAGCGTATCGA 171151
314 pIleAspGlnCys.....SerSerAspGlyTrpPheS 325
171150 CATAAACCCAGTGTCCATTGGGTGAAGCAACAAGACCTTAATCACTTTG 171101
325 erGlyThrHisLysCysHisLeuAsnAsnSerGluCysMetProIleLys 341
171100 CGGATACGCCCGGTGTAAAAAGAACACGACAGAGTGTGAACCATTTACA 171051
342 GlyLeuGlyPheValLeuGlyAlaTyrgluCysIleCysLysAlaGlyPh 358
171050 GGCTGGGSCCTTTAGCGCGGTGGCTACCAGTGCCTGTTGTAAGCCAGGTTT 171001
358 eTyrlHisProGlyValLeuProValAsnAsnPheArgArgGlyProA 375
171000 TCGGCTGCCCAACGTAGTG.....CGGCGACCTTATCTGG 170966
375 spGlnHisIleSerGlySerThrLysAspValSerGluGluAlaTyrlVal 391
170965 CGGAGATTGTGGAGCGCGCATCGCGAACAAGTACTACAACGAGTACGAC 170916
392 CysLeuProCysArgGluGlyCysProPhe.....CysAlaAs 404
170915 TGCCTT.....AAGATTGGCTGTATGTTTAAAGTAGCAATATGTANAAG 170872
404 p.....AspSerProCysPheValGlnGluAspLysTyrlLeuArgL 418
170871 TATGAGATTTGACTCTTGTATGTTT..... 170847
418 euAlaIleLeuSerPheGlnGlyLeuCysMetLeuLeuAspPheValSer 434
170846 .....TTTTTAGGGATCCAAAGCTTCCCATTCAGTGGGATAAGGCCCTCC 170802
435 MetLeuValValTyrlHisPheArgLys 443
170801 .....TACCACATTCCGCCAA 170787
```

seq_name: gb_in2:AE003732

seq_documentation_block:

LOCUS AE003732 234369 bp DNA INV 05-OCT-2000
Drosophila melanogaster genomic scaffold 142000013386035 section 57
DEFINITION of 105, complete sequence.

ACCESSION AE003732 AE002708

VERSION AE003732.1 GI:7300654

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 234369)

AUTHORS

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J., Basu,A.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borokova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S.,

mRNA

gene

```

alignment_scores:
  Quality: 183.50      Length: 511
  Ratio: 0.913        Gaps: 28
  Percent Similarity: 39.335  Percent Identity: 22.701

alignment_block:
US-09-775-181-4 x AE003732/rev  ..

Align seg 1/1 to reverse of: AE003732 from: 1 to: 234369

7  ProLeuLeuLeuCysLeuLeu.....LeuAlaGlnLeuGlyLeuGlyAl 21
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12771 CCCTGCTCATCTCAGAAATATGCCCTTACAGCAAGCTTCTTCGGAGC 12722

21  aValGlyAlaSerArg.....AspProGlnGlyArgProAsps 34
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12721 TACATCTCCAGTCGAGGTATTTCGACCCGCCCACTCAACGACACTACGA 12672

34  erProArgGluArGThrProLysGlyLysProHisAlaGlnGlnProGly 50
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12671 TCCGGCATGATGTACTACTTCTGTCACCGTAGCCGATGATCCG... 12625

51  ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAsp.G 67
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12624 .....CCAACCCACATATCAACGCCCTCG 12602

67  lyThrIleLeuAlaGlnLysLeuAlaGluGluValProMetAspValAla 83
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12601 GCCGTGACTT.....CTCCCCCAACAGCTCGTATTCTGTCGTGTA 12561

84  SerTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsn..... 98
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12560 TCGGGGCTTCTTCAATAAGACGTTCCCGAGATTCCGGCCAAAGACCTTCA 12511

98  ..... 98

12510 GGCTGGGACGACTTCAACGATCCCAATTCATCTGCAGAAGATATCAACGTGG 12461

99  .....CysSerGlyArgTyrGluLeuAlaGlyLeuProGlyLys 111
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12460 AATACTTTCGATGTTCAGGA.....TCTGGGGGCCCATCA... 12426

112  TrpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspTh 128
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12425 .....CCGGGACTCCAT.....ATCCAAGG 12406
  
```

138 rLeu..ThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLys 144
||| |||||| ::||
12405 ACTACACCAGCCTGTATAAAATAAAC..... 12377

145 SerArgGlUGlnAsnLeuGlnAspLeuAspTrpTyrGlnAlaLeuVa 161
:::|||||:::||:
12376GAGTGGTAGCCGCCTCGCT 12357

161 lTrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAlaIleThrP 178
: :::: |||||
12356 ACCAGACAACGTCGAGGA..... 12338

178 heSerThrASPserLeuSerAlaProLaProGlnValPheLeuGlnAla 194
12338 12338

195 ThrArgGlUGlUSerArGileLeuLeuClnAspLeuSerSerAlaPr 211
|||:::~::~:~::~||| |
12337 ..CGGCACGATAGAGATCACTACCAG.....GTGGAAAT 12303

211 oHisLeuAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuAra 228
::: ||||| ||| |||||
12302 CGCTATCGGAACAACAACAGAGACTATACCTCCAC..... 12263

228 rgLysTrpArgProHISLeuHisArgArgGlyPro..... 239
12262GGACCGCTGCTCGAAGA 12242

240 AsnGlnGlyProArGlgLyLeuGlyHisSerTrpargylasaspGlyLe 256
||| |||||
12241 AACCTTGCTCG..... 12230

256 uGlyGlyaspLysSerHisPheLysTrpSerProProfYrLeuGluCySG 273
:::::~::~:~::~|||~::~:~::~|||
12229ATTAAATTACAGGCCGTACTTCGTTGTT. 12200

273 luAsnGlySerTyrlLys.ProGlyTrp..... 281
||| :::: |||||
12199GGCAGGTCCAACAAGTGGTGGTGGCCGAGTAGTCCCAATTGCG 12155

282LeuvalThrLeuSer.....SerA 288
|||~::~:~::~|||~::~:~::~|||
12154 GATATCTACCCCGACACACGAGTCCGTGCACATTGAGTATCCCAAGTA 12105

288 laileTyrgLYLeuGlnProAsnLeuValPro..... 298
||| ~::~:~::~|||~::~:~::~|||
12104 AGATACCTTGAATATCCCTGAATACCTCTCTTTATCTACTGATCGCT 12055

299 ...GlupheArgglyValMetLysValaspIleaSnLeuGlnLysValas 314
~~~~~::|||  
12054 TTTAGATACACGGCGTTTTCGTTCTTGCAGATGGACTCGAGCGTATCGA 12005

314 pIleaspGlnCys.....SerSeraspGlyTrpphes 325  
|||~::~:~::~|||~::~:~::~|||  
12004 CATAAACACAGTGTCCATTGGGTGAAGCAACAAGGACCTAATCACATTG 11955

325 erGlyThrHisLysCysHisLeuaSnAsnSerGliuCysMetProfileLys 341  
::: ||| ~::~:~::~|||~::~:~::~|||  
11954 CGGATACGGCGCGGTGTAAAAAGAACGACGAGAGTGTGAACCATTAACA 11905

342 GlyLeuGlyPheValLeuGlyAlaTyrgLucySileCysLysAlaglyph 358  
|||~::~:~::~|||~::~:~::~|||  
11904 GGCTGGGCGTTAGCGCGGTGGCTACCAGTCCGCTGTAAGCCAGGTTTT 11855

358 eTyrrhisProGlyValLeuProValasnPharGargArgGlyProA 375  
...|||~::~:~::~|||~::~:~::~|||  
11854 TCGGTGTCCCCACGTAGT.....CGGCGACCTTATCTGG 11820

375 spGlnHisIleSerGlySerThrLysaspValSerGliUgluaIatyrVal 391  
~~~::|||~::~:~::~|||~::~:~::~|||  
11819 CCGAGATTGTGGAGCGCGCATCGGCACAACAGTAGTACTACAACGAGTAGCAC 11770

392 CysLeuProCYcArqGLuglyCYCsPropHe.....CysAlaAs 404

```

|||||
11769 TGCCTT.....ANGATTGCTGTATGTTTAAAGTAGCAATATGTAAAG 11726
:
404 p.....AspSerProCysPheValGlnGluAspLysTyrLeuArgL 418
:|||||
11725 TATGAGATTGACTCTTGATGTTT..... 11701
:
418 euAlaIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSer 434
:|||||
11700 .....TTTTAGGATCCAAAGCTTCCCATTCAGTGGGATAAGGCTCC 11656
:
435 MetLeuValTyrHisPheArgLys 443
:|||||
11655 .....TACCACATTCGCCAA 11641

```

seq_name: gb_in3:DMX99675

seq_documentation_block: 3325 bp mRNA INV 07-FEB-1997
 LOCUS DMX99675
 DEFINITION D.melanogaster mRNA for metabotropic glutamate receptor.

VERSION X99675

KEYWORDS metabotropic glutamate receptor.
 fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 3325)

Parmentier, M.L., Pin, J.P., Bockaert, J. and Grau, Y.

Cloning and functional expression of a Drosophila metabotropic
 glutamate receptor expressed in the embryonic CNS
 J. Neurosci. 16 (21), 6687-6694. (1996)

2 (bases 1 to 3325)

Parmentier, M.

Direct Submission

Submitted (31-JUL-1996) M. Parmentier, CCIPE-UPR 9023, Upr 9023,
 141 Rue De La Cardonille, 34094 Montpellier Cedex 05, F-34094,
 FRANCE

FEATURES

source

```

Location/Qualifiers
1. 3325
/organism="Drosophila melanogaster"
/strain="Oregon R"
/db_xref="taxon:7227"
/chromosome="4"
/dev_stage="adult"
190..3120
/codon_start=1
/product="metabotropic glutamate receptor"
/protein_id="CAA67993.1"
/db_xref="GI:1834427"
/db_xref="FLYBASE:FBgn0019985"
/db_xref="SWISS-PROT:P91685"

```

CDS

```

/translaton="MKQKNNGTILVVMVLSWRVVDLKSPTHQDSVSVSLPGD
IILGLPFVHEKGEAGCGPKVYNRGVORLEAMLYADRVNNDNPILPGHIGVHID
TCSRDYVALNQSLQFVQSPASTAKTLSKDEGADSSPOLRNKASSCPVFGVIGSYSSV
SLOVANLRLFLHPQVSPASTAKTLSKDEGADSSPOLRNKASSCPVFGVIGSYSSV
YVSTIHSEGSVGEYGEIHALKEATERNVCIATAEKVPSAADKVFDSILSKLQKPNV
RGVLFTRAEADRILOAKRANLSQPHMTASDGMKQKLLGLELDIAGATVTEL
QSEIIADFRIYMQLTETQNRNFWFAEYEDTNCVLSVAPDTSNSANSTDNI
VKAKTEDDSYRUSEKVGESKTFQVVDVAFAYALHNLHNDRCNTQSDQTTET
RKLQSESVYRKISTDTKSOACPDMDYDGEFNNYLLNVSFIDLAGSEVRFDROG
DGLARYDILYQROENSSGYQYKVIQKFWGLQNLSETVVMNKEQTPSACSLPCSV
GMIKKQDQTCCTCDSEFEYVYDEFTCKDCGGLMPYADKLSCYALDIQYMKWNS
LEFALIPALIFGALT'SIVTLFAKNIDPLVRASGRELSYTLFGILCYCPTFAL
IAKTSICSVLORFIVGVSFIYSALLTYNRIHSKSAQKLTQSPQSV
CTIYAIKRIKPNFNSKTFGTMTYTCIIWLAFAPIYFTGNSYEVQVTLILIS
LSASVALVCLYSPKVIILVFPDKNRKLTMNSTVYRRSAAVAQAQPTSGSYRTHA
PGTSALTGGAGVTNASSSTLPTQNSPHLDEASQTNVAHNTNGEFLPEVGERVEPIC
INVK"

```

BASE COUNT 1038 a 680 c 687 g 920 t

.. .

ORIGIN

alignment_scores:

Quality: 164.00 Length: 494
 Ratio: 0.710 Gaps: 24
 Percent Similarity: 46.761 Percent Identity: 20.648

alignment_block:

US-09-775-181-4 x DMX99675 ..

Align seg 1/1 to: DMX99675 from: 1 to: 3325

```

145 SerArgGluGlnAsnLeuLeuAspLeuAspLysTyrGlnAlaLeuVa 161
:|||||
1072 GCTAAACGAGCTTAACATTATCGCAACCACTTTCATTGGATAGCAAGTGATGG 1121
:|||||
161 lTrp.....SerLeuLeuGluGluProSerIleSerArgA 174
:|||||
1122 TTGGGGTAAACAAACAACTACTCGAAGCTTAGAGGACATCGCTGAGG 1171
:|||||
174 laAlaIleThrPheSerThrAspSer...LeuSerAlaProAlaProGln 189
:|||||
1172 GTGCATATACTAGAACTGCGAGTCAGATATATCGACATTTTGATCGC 1221
:|||||
190 ValPheGlnAlaThrArgGlu.Glu..... 198
:|||||
1222 TACATGATCAACTGACGCCAGACAGACAACTCAAGCAATCTTGGTTGC 1271
:|||||
198 ..... 198
:|||||
1272 TGAATACTGGGAGATACATTAACCTGTGTTTGGACATCTCTTCGGTAA 1321
:|||||
199 .....SerArgIleLeuLeuGlnAspLeuSerSerAl 210
:|||||
1322 AACCTGATACTTCAAACTCGGCTAATTTCTACTGACAAATAAATCGCGTC 1371
:|||||
210 aProHisLeuAlaAsnAlaThrLeuGluThrGluTrpPhe..... 223
:|||||
1372 AAAGCAAAACAGAAATCGCAGCATAGTTACCGGTTTATCTGAAAAAGTTGG 1421
:|||||
224 .....HisGlyLeuArgArgLysTrpArgProHis 233
:|||||
1422 ATACGAAACAGGAGTCAAGACACAATTG.....TTGTGGATGCGCTCT 1465
:|||||
234 LeuHis..... 235
:|||||
1466 AGCATTTGCTTATGCATCTCATAATTTGCATATATGATCGTTGCAACACA 1515
:|||||
236 .....ArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHis..Se 249
:|||||
1516 CAAAGCGACCAACGAGCTGACAGCAAGCAAGCACTACAAGCGAATCTGT 1565
:|||||
249 rTrpArgArgLysAspGlyLeuGlyClyAspLysSerHisPheLysTrpS 266
:|||||
1566 GTGGTATCGAAAA.....ATTTCGACTGAT...ACTAAGTCCCAAGCCT 1606
:|||||
266 ecProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeu 282
:|||||
1607 GTCTCGACATGCTAATATGACGGAAGAGTGTTCACACAAATTACCTT 1656
:|||||
283 ValThrLeuSerSerAla...IleTyrGlyLeuGlnProAsnLeuValPr 298
:|||||
1657 TTGAAGCTATCATTCATACATCTTCTGCGCAGCAAGTAAATTCGATCG 1706
:|||||
298 oGluPheArgGlyValMetLysValAspIle...AsnLeuGlnLysValA 314
:|||||
1707 ACAGGCGCATGTTGGCCAGATACGACATTTTGAATTTATCAACCG.... 1752
:|||||
314 spIleAspGlnCysSerSerAspGly..... 322
:|||||
1753 .....CAAGAAACTCTTCAGGATATCAGTATTAAGGTTATTGAAAA 1794
:|||||

```


OM of: US-09-775-181-4 to: N_Geneseq_0601:* out_format : pfs
Date: Nov 1, 2001 11:04 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODE=frame-p2n.model -DEV=xlh  
-Q/cgn2_1/USPTO_spool/US09775181/runat_30102001_161056_13847/app_query.fasta_1.1955  
-DB=N_Geneseq_0601 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09775181_@CEN1_1_509 -NCPU=6 -ICPU=3 -LONGLOG  
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-775-181-4
Query length: 599
Database: N_Geneseq_0601:*
Database sequences: 730101
Database length: 313950809
Search time (sec): 192.710000

score_list:

| Sequence | Strd Orig | zScore | Escore Len | Documentation |
|--|-----------|--------|------------|---------------|
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAA41583 | | 260.00 | 377.96 | 4.4e-13 |
| /SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ31057 | | 173.50 | 223.51 | 0.0002 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ31058 | | 173.50 | 219.90 | 0.0003 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30616 | | 173.00 | 222.13 | 0.0002 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ30615 | | 171.50 | 219.67 | 0.0003 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ30922 | | 161.00 | 203.71 | 0.0022 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ31049 | | 161.00 | 203.71 | 0.0022 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ51398 | | 161.00 | 203.71 | 0.0022 |
| /SIDS1/gcgdata/geneseq/NA1996.DAT:AAZ61382 | | 161.00 | 201.92 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ6964 | | 161.00 | 201.92 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ5859 | | 161.00 | 201.92 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ25055 | | 161.00 | 201.92 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ82485 | | 161.00 | 201.92 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ89298 | | 161.00 | 201.92 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ50617 | | 158.50 | 202.06 | 0.0028 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ6962 | | 152.50 | 186.11 | 0.0214 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ5857 | | 152.50 | 186.11 | 0.0214 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ25053 | | 152.50 | 186.11 | 0.0214 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ82483 | | 152.50 | 186.11 | 0.0214 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ89296 | | 152.50 | 186.11 | 0.0214 |
| /SIDS1/gcgdata/geneseq/NA1996.DAT:AAZ61381 | | 150.00 | 183.07 | 0.0316 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ6963 | | 150.00 | 183.07 | 0.0316 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ5858 | | 150.00 | 183.07 | 0.0316 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ25054 | | 150.00 | 183.07 | 0.0316 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ82484 | | 150.00 | 183.07 | 0.0316 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ89297 | | 150.00 | 183.07 | 0.0316 |
| /SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ86166 | | 148.50 | 183.35 | 0.0305 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ6965 | | 143.50 | 175.78 | 0.0805 |
| /SIDS1/gcgdata/geneseq/NA1998.DAT:AAZ5860 | | 143.50 | 175.78 | 0.0805 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ25056 | | 143.50 | 175.78 | 0.0805 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ82486 | | 143.50 | 175.78 | 0.0805 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ89299 | | 143.50 | 175.78 | 0.0805 |
| /SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ89290 | | 139.50 | 170.01 | 0.1688 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ1435 | | 135.00 | 127.00 | 41.99 |
| /SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ30887 | | 134.00 | 166.16 | 0.3573 |
| /SIDS1/gcgdata/geneseq/NA1995.DAT:AAQ80418 | | 131.00 | 158.32 | 0.7564 |
| /SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ86165 | | 130.00 | 158.47 | 0.7413 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ1451 | | 129.50 | 128.76 | 33.51 |
| /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ50743 | | 127.50 | 157.05 | 0.8904 |
| /SIDS1/gcgdata/geneseq/NA1992.DAT:AAQ5815 | | 126.50 | 148.66 | 2.61 |

/SIDS1/gcgdata/geneseq/NA2000.DAT:AAA59555 + 125.50 154.71 1.20
/SIDS1/gcgdata/geneseq/NA1997.DAT:AAZ86167 + 123.50 149.64 2.30
/SIDS1/gcgdata/geneseq/NA1992.DAT:AAQ25812 + 122.50 145.04 4.15
/SIDS1/gcgdata/geneseq/NA1999.DAT:AAZ31055 + 120.00 146.85 3.29
/SIDS1/gcgdata/geneseq/NA1996.DAT:AAZ16710 + 120.00 146.84 3.30

seq_name: /SIDS1/gcgdata/geneseq/NA2000.DAT:AAA41583

seq_documentation_block:

ID AAA41583 standard; cDNA; 182 BP.
XX
AC AAA41583;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:323.
XX
XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
XX antitumor; osteoprotective; neuroprotective; antipressant; antipruritic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX Homo sapiens.
OS
XX WO2000021990-A1.
PN
XX 20-APR-2000.
PD
XX 15-OCT-1999; 99WO-US24205.
PF
XX 11-OCT-1998; 98US-0104435.
PR
XX (CEMY) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Morberg D, Treacy M;
XX Whitt; 2000-317937/27.
DR
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (SESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX Claim 1; Page 243; 618pp; English.
XX
XX AAA41261 to AAA43419 represent specifically claimed secreted expressed
XX sequence tags (SESTs), isolated from human, mouse, xenopus and rat
XX tissue sources. The SESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnary; antitumor; osteoprotective; neuroprotective;
XX neuroprotective; antiparkinsonian; antipruritic; cerebroprotective;
XX anticonvulsant; and antidepressant. The SESTs can be used for gene
XX therapy and in vaccines. The SESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions
XX (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
XX osteoporosis, osteoarthritis, central nervous system disorders

CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 182 BP; 46 A; 36 C; 49 G; 51 T; 0 other;

alignment_scores:
Quality: 260.00 Length: 51
Ratio: 5.200 Gaps: 0
Percent Similarity: 98.039 Percent Identity: 98.039

alignment_block:
US-09-775-181-4 x AAA41583 ..

Align seg 1/1 to: AAA41583 from: 1 to: 182

518 IleProTyrMetThrGlyArgValMetArgMetLeuAlaValIleLe 534

|||||
27 ATTCATATATGACTGGCGGACGGTTCATGAGGATGCTGGCAGTAATACT 76

534 uLeuValValPheTrpPheLeuIleGlyTyrThrSerSerValCysGlnA 551

|||||
77 CTTGGTGTGTGTGTGTTCATTTGGCTGGACTTCATCTGTGTGCCAGA 126

551 snLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHis 567

|||||
127 ATTTGGAGAACAGATTTCATTATTGGCCAGGGAACACCCGATCAC 176

568 Leu 568

||||
177 CTC 179

seq_name: /STD51/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31057

seq_documentation_block:

ID AAZ31057 standard; cDNA; 2751 BP.

XX AAZ31057;

XX 07-JAN-2000 (first entry)

XX Chimeric hCAR/hmGluR2 coding sequence.

XX G-protein fusion receptor; CaR; calcium receptor; GluR; head injury;
XX metabotropic glutamate receptor; GABAR; chimeric receptor; stroke;
XX gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
XX spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
XX Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
XX cognitive disorder; ss.

XX Homo sapiens.

XX WO9951641-A1.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSF-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;
PI Simin RT;

XX WPI; 1999-610995/52.

XX P-PSDB; AAY49126.

XX New G-protein fusion receptors and chimeras containing domains from
PT different receptors, used to screen for modulators, potentially useful

PT e.g. for treating or preventing stroke or Alzheimer's disease -
XX Disclosure; Fig 9; 255pp; English.
XX The invention relates to G-protein fusion receptors (I) comprising:
CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
CC intracellular (ICD) domains, each chosen independently from a CaR
CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABAR
CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
CC linker. (I), and recombinant chimeric receptors (CR) without the GP
CC component, are used to assess function of the various domains and to
CC identify compounds (e.g. allosteric modulators or antagonists) that act
CC on these domains. The modulators are potentially useful for treating or
CC preventing diseases associated with the receptors, e.g. stroke, head or
CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
CC (1) for recombinant production of corresponding proteins; and (2) to
CC produce cells used in screening for modulators. Use of CaR and mGluR
CC domains allows presentation of GABAR domains, to a binding agent, in a
CC form more like the natural domain structure compared with use of
CC incomplete receptors, lacking one or more domains. By shuffling different
CC domains, agents can be identified that affect particular domains of a
XX receptor.
SQ Sequence 2751 BP; 601 A; 772 C; 735 G; 643 T; 0 other;

alignment_scores:

Quality: 173.50 Length: 641
Ratio: 0.655 Gaps: 29
Percent Similarity: 41.342 Percent Identity: 20.749

alignment_block:

US-09-775-181-4 x AAZ31057 ..

Align seg 1/1 to: AAZ31057 from: 1 to: 2751

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91

|||||
394 TCAGACACATTCCTCTAGCATGTGCTGGGAGCAACTGGCTCAGG 443

91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108

|||||
444 CGTCTCCAGCGCAGTGCACAAATCTGCTGGGGCTCTTACATT..... 486

108 euProGlyLysTyrProAlaLeuAlaSerAlaHisProSerLeuHisArg 124

|||||
487CCCCAGGTCACTTATGCCCTCCTCCAGC.....AGA 516

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141

|||||
517 CTCTCAGCAACAAGAAATCAATTCAGTCTTCTCCGAAACC...ATCCC 563

141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTyrTrp 158

|||||
564 CAATGATGACACCGCCACTGCCATGGCAGACATCATCGAGTATTTC 613

158 ln.AlaLeuValTrpSerLeuLeuGlyGluProSerIleSer..... 172

|||||
614 GCTGGAACCTGGTGGGCACAATTCACGCTCATCGACGACTATGGCGGCCG 663

173ArgAlaAlaIleThrPheSerT 180

|||||
664 GGATTTGAGAAATTCGAGAGGAAGCTGAGAAAGGATATCTGCATCGA 713

180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195

|||||
714 CTTCACTGAACATCTCCAGTACTCTGTGATGAGGAGAGATCCAGCATG 763

196ArgGluGluSer.ArgIleLeuLeuG 204

|||||
||| :||| ||| :|||

```

764 TGGTAGAGGTGATTCAAATTCACGGCCAAAGATCATCTGTTTCTCC 813
204 lAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
814 AGTGGCCAGATCTTGAGCCCTCATCAAGAGAGATTGTCGGCGCAA... 860
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
861 .....TATCAGCGGCAAGATCTGGCTGCCAGCGAGCCTGGCGCAGCT 904
237 gGlyProAsnGly.....ProArgGlyLeuGlyHisSerT 250
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
905 CCTCCTGATCGCCATCGCTCAGTACTTCCACGTGTGGCGGCACCAT 954
250 rPArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
955 GGATTGCTCTGAAGC.....TGGGCA 977
267 ProProTyLeuGlu.....Cy 272
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
978 GATCCCGAGGCTTCGGGAATTCCTGAAGAGTCCATCCAGGAAGTCTG 1027
272 s.GluAsnGlySerTyLysProGlyTrpLeuValThr..... 284
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1028 TCCACAATGTTTGGCAGAGGATTGGGAAGACATTTAACTCCAC 1077
285 LeuSerSerAlaLeuTyGrGlyLeuGlnPro..... 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1078 CTCGAAGAAGGTGCANAAGAGCCTTTACCTGTGGACACCTTTCTGAGAGG 1127
295 .....AsnLeuValProGluPheArg 302
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1128 TCACGAGAAAGTGGCGAGGTTTAGCACACAGCTGCAGAGCCTTCGCAC 1177
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1178 CCCTGTGTACAGGGATGAGACATCAGCAGTCTCAGAGCCCTTACATA 1227
313 ..... 313
1228 GATTACACGATTTAGGGATACCTACAATGTGTACTTAGCAGTCTACTC 1277
314 .....AspIleAspGlnCys..... 318
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1278 CATGCCCCAGCCTTGCAAGATATATACCTGCTTACCTGGGAGAGGC 1327
319 .....SerSerAspGlyTrpPhe 324
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1328 TCTTACCAATGGCTCTCTGCAGACATCAAGAAAGTTGAGCGGTGGCAG 1377
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1378 GTCTGAAGCACCTACGGCATCTAACTTTACAAACAATATGGGGAGCA 1427
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1428 GGTGACCTTTGATGAGTGTGGTGAACCTGTGGGAACTATTCCATCATCA 1477
348 .....Gly 348
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1478 ACTGGCACCTCTCCCGAGAGTGGCTCCATCGTGTTTAAGGAAGTCGGG 1527
349 AlaTyrgluCysIleCysLysAlaGly.....PheTyHisProGly 362
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1528 TATTACAGCTGTATGCCAAGAAGGAGAAAGACTCTTTCATCAACGAGGA 1577
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1578 GAAATCTGTGTGAGTGGTGTCTCCAGGAGGTGCGCTTCTCCAACTGCA 1627
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1628 GCGGAGACTGCTTGGCAGGAGCAGGAAGAGGATCATTTGAGGGGGAGGCC 1677

```

```

389 .....AlaTyrvAlCysLeuProCysArgGluGly..... 398
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1678 ACCTGCTGTTTGGAGTGTGGAGTCTCTGATGGGAGATAGTAGTATGA 1727
399 .....CysProPheCysAlaAsp..... 405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1728 GACAGATGCCAGTGGCTGTAAACAAGTCCCGCATGACTTCTGTGTCAAATG 1777
406 .....SerProCysPheValGlnGluAspLysTyLeuArg..... 417
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1778 AGAACACACACCTCTCTCGAACTGCCCGCAGGAGTACATCCGCTGGGC 1827
418 .....LeuAlaIleLeuSerPheGlnGlyLeuCy 427
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1828 GATGCTGGCTGTGGACCTGTCAACATCGCTGCTCGGTGCCCTGCGC 1877
427 sMetLeuLeuAspPheValSerMetLeuValValTyHisPheArgLysA 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1878 CACCTG.....TTGTGCTGGGTCTTTGTGGGCAC...AATGCCA 1918
444 lAlysSerIleArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPhe 460
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1919 CACCACTGCTCAAGCCTCAGCTCGGACCTCTGCTACATCTCTGCTGGT 1968
461 GlySerLeuLeuLeuTyPheProValValIleLeuTyPheGluProSe 477
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1969 GGTGTCTCTCTGCTACTGATGACCTTCATCTTCTATTGCCAAGCATC 2018
477 rThrPheArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaT 494
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2019 CACGGCAGTGTACCTTACGGGCTTTGGTTGGGCACCTGCTCTCTG 2068
494 hrValTyrglyThrValThrLeuLysLeuHisArgValLeuLysValPhe 510
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2069 TCTGTACTACGCCCTGCTCACCAGACCAACCGCATTCACGACATCTTC 2118
511 .....LeuSerArgThrAlaGlnArgIleProTyMetThr...GlyGI 524
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2119 GGTGGGCGCGGGAGGTGCCAGCGCCAGCGCTTCATCAGTCTGCTC 2168
524 yArgValMetArgMetLeuAlaValIle.....LeuLeuValV 537
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2169 ACAGGTGGCCATCTGCTGGCATTATCTCGGGCCAGCTGCTCATGCTGG 2218
537 alPheTrpPheLeuIle 542
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2219 TCGCCTGGCTGGTGGTG 2235
seq_name: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ31058
seq_documentation_block:
ID_AAZ31058_standard; cDNA; 3831 BP.
XX
AC_AAZ31058;
XX
DT_07-JAN-2000 (first entry)
XX
DE_pHcAr/hmGluR2*Gqi5 fusion construct cDNA sequence.
XX
KW_G-protein fusion receptor; Car; calcium receptor; GluR; head injury;
KW_metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
KW_gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
KW_spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
KW_Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
KW_cognitive disorder; ss.
XX
OS_Homo sapiens.
XX
PN_W09951641-A1.
XX
PD_14-OCT-1999.
XX

```



```

1478 ACTGGCAGCTCTCCCGAGAGGAGTGGCTCCATCGTGTGTTAAGGAAGTCGGG 1527
||||
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGlu 362
||||: ||||| ||| |||
1528 TATTACACAGCTATATGCCAAGAAGGAGAAAGACTCTTTCATCAACGAGGA 1577
362 yValLeuProValAsnAspPheArgArgGlyProAspGlnHis..... 377
||||: ||||| ||||| ||||| |||||
1578 GAAATCTGCTGGAGTGGTGTCTCCAGGAGGAGTGCCTTCTCCAACTGCA 1627
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
||||: ||||| ||||| ||||| |||||
1628 GCGGAGACTGCTGGCAGGAGCAGCAAGAGGATCATTTGAGGGGAGCCCC 1677
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
||||: ||||| ||||| ||||| |||||
1678 ACCTGCTGCTTTTCAGTGTCTGAGTGTCTCTGATGGGAGTATAGTGATGA 1727
399 .....CysProPheCysAlaAspAsp..... 405
|||| ||| |||||
1728 GACAGATGCCAGTGGCTGTGAACAAGTGCAGATGACTTCTGTCTCAATG 1777
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
|||| ||||| ||||| ||||| |||||
1778 AGAACACACCTCTGCTTCGAACTGCCAGGAGTACATCCGCTGGGGGC 1827
418 .....LeuAlaIleIleSerPheGlnGlyLeuCy 427
||||: ||||| ||||| ||||| |||||
1828 GATGCTGGCTGGTGGGACCTGTCAACATCGCTGCTGGTGGCTGGC 1877
427 sMetLeuLeuAspPheValSerMetLeuValValTyrHisPheArgLysA 444
||| ||||| ||||| ||||| ||||| |||||
1878 CACCCCTG.....TTGTGCTGGGTCTTTGTGGGCAC...AATGCCA 1918
444 lAlsSerIleArgAlaSerGlyLeuIleLeuGluThrIleLeuPhe 460
||| ||||| ||||| ||||| |||||
1919 CACCACTGTCAAGGCTCAGTGGGAGCTCTGTACATCTCTGCTGGGT 1968
461 GlySerLeuLeuTyrPheProValIleLeuTyrPheGluProse 477
||| ||||| ||||| ||||| ||||| |||||
1969 GGTGTCTCTCTGCTACTGATGATGACCTTCATCTTATTCACCAAGCAATC 2018
477 rThrPheArgCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaT 494
|||| ||||| ||||| ||||| ||||| |||||
2019 CACGCGAGTGTACCTTACCGGCTCTTGGTTGGCACTGGCTTCTCTG 2068
494 hrValTyrGlyThrValThrLeuHisArgValLeuLysValPhe 510
||| ||||| ||||| ||||| ||||| |||||
2069 TCTGCTACTCAGCCCTGCTCACCAGACCCAGCCGATTCACGCACTTC 2118
511 .....LeuSerArgThrAlaGlnArgIleProTyrMetThr...GlyG 524
||| ||||| ||||| ||||| ||||| |||||
2119 GTGGGGCCGGAGGCTGCCAGCGCCACGCTTCATCAGTCTGCTGCTC 2168
524 yArgValMetArgMetLeuAlaValIle.....LeuLeuValV 537
||||: ||||| ||||| ||||| ||||| |||||
2169 ACAGGTGGCACTCTGCTGGCACTTATCTCGGCCAGCTGCTCATCGTGG 2218
537 aPheTrpPheLeuIle 542
||| ||||| ||||| ||||| ||||| |||||
2219 TCGCCCTGGCTGGTGGG 2235

```

seq_name: /SIDSL/fcgdata/geneseq/geneseqn/NA2000.DAT:AAZ50616

seq_documentation_block:

ID AAZ50616 standard; cDNA; 2922 BP.

XX AC AAZ50616;

XX AC AAZ50616;

DT 20-JUN-2000 (first entry)

DE Human calcium sensing receptor isoform, CasRd encoding cDNA.

```

XX Calcium Sensing Receptor; CasRd; isoform: human; splice variant;
KW hypotensive; osteopathic; receptor activity; calcium level; modulator;
KW treatment; hyperparathyroidism; osteoporosis; Paget's disease;
KW hypercalcaemia malignancy; hypertension; gene therapy; ss.
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
CDS 1..2922
FT /*tag= a
FT /product= "Human calcium sensing receptor isoform, CasRd"
FT /note= "Does not include stop codon"
FT /partial
XX
XX WO200006601-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17116.
XX
XX 30-JUL-1998; 98US-0094702.
XX
XX (AVET ) AVENTIS PHARM PROD INC.
XX
XX Yu KT, Labaudiniere RF, Thrower LW;
XX
XX WPI: 2000-195263/17.
XX
XX P-PSDB; AAY45000.
XX
XX Nucleic acids encoding isoforms of human calcium sensing receptor for
XX treating, e.g. hyperparathyroidism or osteoporosis
XX
XX Claim 8; Page 60-66; 81pp; English.
XX
XX The present sequence is the cDNA encoding an isoform of the human
XX calcium sensing receptor CasRd, expressed in the kidney. It is a splice
XX variant of the wild type CasR, that arise from deletion of nucleotides
XX 1080-1391, from the extracellular domain, that comprise acidic residues.
XX It has hypotensive and osteopathic activity. The CasR isoforms can be
XX used to identify agonists and antagonists that modulate the receptor
XX activity and calcium levels. These modulators are useful for treating
XX hyperparathyroidism or osteoporosis, Paget's disease, hypercalcaemia
XX malignancy or hypertension. The DNA sequence
XX the CasR activity and in gene therapy.
XX Note: This sequence has deletion of nucleotides 1080-1391, according to
XX the sequence shown in the specification, but has been stated as
XX 1075-1386 in the claims.
XX
XX Sequence 2922 BP; 655 A; 848 C; 752 G; 667 T; 0 other;

alignment_scores:
Quality: 173.00 Length: 590
Ratio: 0.634 Gaps: 27
Percent Similarity: 46.271 Percent Identity: 21.525

alignment_block:
US-09-775-181-4 x AAZ50616 ..
Align seg 1/1 to: AAZ50616 from: 1 to: 2922

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
||||: ||||| ||||| ||||| ||||| |||||
394 TCAGAGCACATTCCTCTACGATTCCTGTGGGAGCAACTGGCTCAGG 443
||||: ||||| ||||| ||||| ||||| |||||
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
||| ||||| ||||| ||||| ||||| |||||
444 CGTCTCCAGCGGAGTGGCAATCTGCTGGGCTCTTCTACATT..... 486
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
||| ||||| ||||| ||||| ||||| |||||

```

487CCCAGGTCAGTTATGGCTCTCTCCAGC.....AGA 516

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuGl 141
||||| : : : : : ||||| : : : : : : :
517 CTCCTCAGCAACAAGAAATCAATTCAAGTCTTTCTCCGAACC...ATCCC 563

141 nSerAsnLysSerArgGluInAsnLeuGlnValPheLeuAspTrpTyrG 158
||| : : : : : ||| : : : : : ||| : : : : :
564 CAATGATGAGCACACCGGCATGCCATGGCAGACATCATCGAGTATTTC 613

158 InAlaLeuValTrpSerLeuLeuCluclyGluProSerIleSer..... 172
||||| : : : : : ||||| : : : : : : :
614 GCTGGAAGTGGGTGGCAATAATGCAGCTGATGACGATATGCGGCGCG 663

173ArgAlaAlaIleThrPheSerT 180
||| : : : : : ||| : : : : : ||| : : : : :
664 GGGATTGAGAAATCCGAGAGGAAGCTGAGAAAGGGATATCTCCATCGA 713

180 hrAspSerLeuSerAlaProLaProGlnValPheLeuGlnAlaThr.... 195
||| : : : : : ||||| : : : : : ||| : : : : :
714 CTTTGACTGAACCTCATCTCCAGTACTCTGATGAGAGAGATCCAGCATG 763

196ArgGluGluSer.ArgIleLeuLeuG 204
||| : : : : : ||||| : : : : : ||| : : : : :
764 TGGTAGAGGTGATCAAAAATCCACGGCCAAGTCACTGGTGGTTTTCTCC 813

204 InAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
|| | ||| ||||| ||||| : : : : : : :
814 AGTGGCCAGATCTTGACCCCTCATCAGAGAGATGTCCGGCCCAA... 860

221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
||| ||||| ||||| : : : : : ||| : : : : :
861TATCAGGGCAAGATCTGGCTGGCCAGCAGGCGCTGGGCCAGCT 904

237 gCyProasnGlnGly.....ProArgGlyLeuGlyHisSerT 250
||| : : : : : ||||| : : : : : ||| : : : : :
905 COTCCCTGATGCCATGTGCTCAGTACTTCCACGTGGTGGCGCACCATTT 954

250 tpArgLysAspClyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
|| ||||| : : : : : ||||| : : : : : ||||| : : : : :
955 GGATTGGCTCTGAAGGC.....TGGGCA 977

267 ProPrtYrLeuLeu.....Cy 272
||| ||| ||| : : : : : : :
978 GATCCACGGCTTCGGGAATTCCTGAAGAGGTCCATCCCGAAGAGCTG 1027

272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThrLeuSerSerAla 288
|| : : : : : ||| : : : : : ||||| : : : : :
1028 TCCACATGGTTTTGCCAAGAGGATTTTGGGAAGAACAATTAATTCGCCAC 1077

289 IleTyrGlyLeuGlnProAsnLeuValProclupheaArgGlyValMetLy 305
: : : ||| : : : : : : : : : : : : : ||| : : : : :
1078 CTACGGCATCTA.....AACTTTACAACAATATGGG 1109

305 sValAspIleAsnLeuGlnLysValAspIleAspGlnCysSerSerAspG 322
: : : : : ||||| : : : : : ||||| : : : : :
1110 G.....GAGCAGGTGACCTTTGATGAGTGT.....G 1135

322 lyTrpPheSerGlyThrHisLysCys.....HisLeuAsnAsnSer 335
|| : : : ||||| : : : : : ||||| : : : : :
1136 GTGACCTGTGGGNAACATATTCCATCATCAACTGGCACCTCTCCCCAG 1185

336 Glu...CysMetProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGI 351
: : : : : ||||| : : : : : ||||| : : : : :
1186 GATGCTCCATCGTGTAAAGGAAGTCCGGTAT.....TACA 1223

351 uCysIleCysLysAlaGly.....PheTyrHisProGlyValLeu 365
: : : ||||| ||| : : : : : ||| : : : : :
1224 CGTCTATCCCAAGGAGGAGAAAGACTTTCATCAACGAGGAGAAATCC 1273

365 roValAsnAsnPheArgArgGlyProAspGlnHis..... 377

1274 TGTGGAGTGGTTTCTCCAGAGGAGTGCCCTTCTCCAACTGCAGCCGAGAC 1323

378 ...IleSerGlySerThrLysAspValSerIleGlu...A1 389
1324 TGCCTGGCAGGACCAGGAAAGGATCATTTAGGGGGAGCCACCTGCTG 1373
389 aTyValCysLeuProCysArgGluGly... 398
1374 CTTTGAGTGTGGAGTGTCTGTATGGGGAGTATAGTGTAGAGACAGATG 1423
399CysProPheCysAlaAsp... 405
1424 CAGTGCCTGTAAACAAGTGCCACAGATGACTTCTGGTCCAATGAGAACCAC 1473
406 SerProCysPheValGlnGluAspLysTyrrLeuArg... 417
1474 ACCTCTGCATTTGCCAAGGATGACGATTTCTGTCGGGAGCGAGCCCTT 1523
418 ...LeuAlaIleLeuSerPheGlnGlyLeuCysMetLeuLeuAspPheV 433
1524 TGGGATGCACTACCCCTCTTTGGCGTGCTGGGCATTTTCTGCAGCCT 1573
433 alSerMetLeuValValTyrrHisPheArgLysAlaLysSerIleArgAla 449
1574 TTGTGCTGGTGTCTTTCAAGTTCGCGCACACACACCATTTGCAAGGCC 1623
450 SerGlyLeuLeuLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTy 466
1624 ACCAACCGAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1673
466 rPheProValValIleLeuTyrrPheGluProSerThrPheArgCysIleL 483
1674 CTCAGCTCCCTGTTCTTCATCGGGAGCCAGGACTGGAGCTGCGGCC 1723
483 euLeuArgTrrAlaArgLeuLeuGlyPheAlaThrValTyrrGlyThrVal 499
1724 TGGCGCAGCGGCCCTTTGGCATCAGTTCGTGCTGCTGCATCTCATGCATC 1773
500 ThrLeuLysLeuHisArgValLeuLysValPheLeuSerArgThrAlaGl 516
1774 CTGGTGAACAACACCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1823
516 nArgIleProTyrrMetThrGlyGlyArgValMetArgMetLeuAlaVal. 532
1824 CAGCTTCCACGCAAGTGGTGGGGGCTCAACCTGCAAGTTCCTCTCTGTTT 1873
533IleLeuLeuValValPheTrrPheLeuIle 542
1874 TCCTCTGCACCTTCATGCAGATTTGTCATCTGTGTATCTGGCTACACC 1923
543 GlyTrrPthrSerVal...CysGlnAsnLeuGluLysGlnIleSerLe 558
1924 CGCGCCCCCTCAAGCTACCGCAACACAGAGCTGGAGGATGAGATCATCTT 1973
558 uIle.....GlyGlnGlyLysThrSerAspHisLeuIlePheA 571
1974 CATCAGTGCACAGGGCTCCCTCATGGCCCTGGGCTTCTCTGCGCT 2023
571 snMetCysLeuIle 575
2024 ACACCTGCTGCTG 2037

```
seq_name: /SIDS1/gcgdata/geneseq/NA2000.DAT:AAZ50615
seq_documentation_block:
ID   AAZ50615 standard; cDNA; 3003 bp.
XX
XX
AC   AAZ50615;
XX
XX
DT   20-JUN-2000 (first entry)
XX
XX
DE   Human calcium sensing receptor isoform, CaSRc encoding cDNA.
XX
XX
KW   Calcium Sensing Receptor; CaSRc; isoform; human; splice variant;
```


hypotensive; osteopathic; receptor activity; calcium level; modulator;
treatment; hyperparathyroidism; osteoporosis; Paget's disease;
hypercalcaemia malignancy; hypertension; gene therapy; ss.

XX Homo sapiens.
XX Synthetic.

PH Key Location/Qualifiers
XX CDS 1..3003
FT FT /*tag= a
FT FT /product= "Human calcium sensing receptor isoform, CaSRc"
FT FT /note= "Does not include stop codon"
FT FT /partial

XX W0200006601-AL.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17116.

XX 30-JUL-1998; 98US-0094702.

XX (AVET) AVENTIS PHARM PROD INC.

XX Yu KT, Labaudiniere RF, Thrower LW;

XX WPI; 2000-195263/17.

XX P-PSDB; AAY44999.

XX Nucleic acids encoding isoforms of human calcium sensing receptor for
treatment, e.g. hyperparathyroidism or osteoporosis -

XX Claim 8; Page 50-56; 81pp; English.

XX The present sequence is the cDNA encoding an isoform of the human
calcium sensing receptor CaSRc, expressed in the kidney. It is a splice
variant of the wild type CaSR, that arise from deletion of nucleotides
1380-1610, from the extracellular domain, that comprise acidic residues.
It has hypotensive and osteoplastic activity. The CaSR isoforms can be
used to identify agonists and antagonists that modulate the receptor
activity and calcium levels. These modulators are useful for treating
hyperparathyroidism or osteoporosis. Paget's disease, hypercalcaemia
malignancy or hypertension. The DNA sequence is also useful for altering
the CaSR activity and in gene therapy.

XX Note: This sequence has deletion of nucleotides 1380-1610, according to
the sequence shown in the specification, but has been stated as
1378-1608 in the claims.

XX Sequence 3003 BP; 677 A; 880 C; 763 G; 683 T; 0 other;

alignment_scores:
Quality: 171.50 Length: 619
Ratio: 0.638 Caps: 27
Percent Similarity: 43.457 Percent Identity: 21.486

alignment_block:

US-09-775-181-4 x AAZ50615 ..

Align seg 1/1 to: AAZ50615 from: 1 to: 3003

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
: :::::::::::::: :::::
394 TCAGAGCACATTCCCTCTACGATTGCTGGTGGGAGCAACTGGCTCAGG 443

91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
: :::::::::::::: :::::
444 CGCTCCACGGCAGTGGCAAAATCTCTGGGCTCTTACATT..... 486

108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
: :::::::::::::: :::::
487CCACAGGTCAGTTATGCTCTCTCCAGC.....AGA 516

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
: :::::::::::::: :::::
517 CTCCTCAGCACACAGAATCAATTCAAGTCTTTCTCCGGAACC...ATCCC 563
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
: :::::::::::::: :::::
564 CAATGATGAGCACAGGCCACTGCCATGGCAGACATCATCGAGTATTTCC 613
158 InAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
: :::::::::::::: :::::
614 GCTGGAACATGGGTGGGCACAAATTCAGCTGATGACGACTATGGCGGCG 663
173ArgAlaAlaIleThrPheSerT 180
664 GGGATTGAGAAATCCGAGAGGAAGCTGAGGAAGGGATATCTGCATCGA 713
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
714 CTTCACTGAATCATCTCCAGTACTCTGATGAGGAAGAGATCCAGCATG 763
196ArgGluGluSer.ArgIleLeuLeuG 204
764 TGGTAGAGGTGATTCAAAATTCACGGCCAAAGTCACTCGTGGTTTCTCC 813
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
814 AGTGGCCAGATCTTTGAGCCCTCATCAAGGAGATTTGCCGCGCAA... 860
221 GluTrpPheHisGlyLeuArgLysTrpArgProHisLeuHisArgAr 237
861TATCACGGCAAGATCTGGCTGCCAGCGAGGCTCTGGCCAGCT 904
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
905 CTTCCCTGATGCCATCCCTCAGTACTTCCAGCTGGTTGGCGGCACCAT 954
250 tpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
955 GGATTTCGCTCTGAAGGC.....TGGGCA 977
267 ProProTyrLeuGlu.....Cy 272
978 GATCCACAGCTTCCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1027
272 s.GluAsnGlySerTyrLysProGlyTyrLeuValThr..... 284
1028 TCCACAAATGGTTTCCCAAGGAGTTTGGGAAGAACAATTTAACTGCCAC 1077
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
1078 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTGAGAG 1127
295AsnLeuValProGluPheArgG 302
1128 TCACGAAGAAGTGGCGACAGGTTTAGCAACAGCTCGACAGCCTTCCGAC 1177
302 lyValMetLysValAspIleAsnLeuGlnLysValAsp.....Ile 315
1178 CCTCTGTACAGGGGATGAGACATCAGAGTGTGAGAGCCCTTACATA 1227
316 AspGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysHisLe 332
1228 GATTAC.....ACGCAATTT 1241
332 uAsnAsnSer..... 335
1241 ACAGGATATCTTACAATGTACTTAGCAGTCTACTCCATTTGCCACGCT 1291
336GluCysMetProIleLysGlyLeuGlyPheValLeu 347
1292 TCACAGATATATATACCTGCTTACCTGGGAGAGGCTC...TTCACCAAT 1338
348 GlyAlaTyrGluCysIleCysLysAlaGlyPheTyrHisProGlyValLe 364

```
|||||
1339 GCCTCTGTCGACAGCAAGAAAGTTGAGCGGTGGCAG.....GT 1379
364 uProValAsnAsnPheArgArgArgGlyProAspGlnHisIleSerGlyS 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1380 GCCCTTCTCAACTGCAGCCGA.....GACTGCTGCGCAGGA 1417
381 erThrLysAspValSerGluGlu.....AlaTyrValCysLeu 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1418 CCAGAAAGGATCATTTGAGGGGAGCCACCTGCTGCTTTGAGTGTGTG 1467
394 ProCysArgGluGly.....CysPr 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1468 GAGTGTCTGATGGGAGTAGTAGTGAGACAGATGCCAGTGCCTGTAA 1517
400 oPheCysAlaAspAsp.....SerProCysPheV 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1518 CAAGTGCACAGATGACTTCTGGTCCANTGAGAACACACCTCTGTCATG 1567
410 aGlnGluAspLysTyrLeuArg.....LeuAlaIle 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1568 CCAAGGAGATCGAGTTCTGCTGGGAGCGCCCTTTGGGATCGCACTC 1617
421 IleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuVa 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1618 ACCCTCTTTCGCGTGGCGATTTCTGACAGCCCTTTGTCGTGGTGT 1667
437 lValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleL 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1668 GTTATCATCAAGTTCGCAACACACCAATGTCAGGCCACCAACCGAGCG 1717
454 euLeuGluThrIleLeuPheGlySerLeuLeuTyrPheProValVal 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1718 TCTCTACCTCCTCTCTCTGCTGCTGCTGCTCTCTCTCTCTCTCTCT 1767
471 IleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAl 487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1768 TTCCTCATCGGGAGAGCCCAAGGACTGGACGTGCGCGCTGCGCCAGCGCG 1817
487 aArgLeuGlyPheAlaThrValTyrGlyThrValThrLysLeuH 504
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1818 CTTTGGCATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1867
504 isArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyr 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1868 ACCGTGTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
521 MetThrGlyArgValMetArgMetLeuAlaVal..... 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1918 AAGTGTGGGGGCTCAACCGCAGTTCCTGCTGCTGCTGCTGCTGCTGCT 1967
533 .....IleLeuValValPheTrpPheLeuIleGlyTrpThrSers 547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1968 CATGCAGATGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2017
547 erVal...CysGlnAsnLeuGluLysGlnIleSerLeuIle..... 559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2018 GCTACCCCAACAGGAGGATGAGATCATCTTCATCAGGTGCGCAC 2067
560 ...GlyGlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeuI 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2068 GAGGGTCCCTCATGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117
575 e 575
2118 G 2118
```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:AA90922

seq_documentation_block:

ID AA90922 standard; DNA; 3234 BP.

XX

AC AA90922;

```
XX 17-JAN-2000 (first entry)
XX Human calcium receptor nucleic acid sequence.
DE Human calcium receptor nucleic acid sequence.
XX Human calcium receptor; GABAB modulator; GABAB receptor protein;
KW gamma-aminobutyric acid receptor; metabotropic receptor;
KW chimeric protein; synaptic transmission; extracellular binding domain;
KW antagonist activity; intracellular domain; intracellular effect; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..3234
FT /tag= a
FT /product= "Human calcium receptor protein"
XX
PN WO9951636-A2.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-US07352.
XX
PR 03-APR-1998; 98US-0080676.
XX (NPSP-) NPS PHARM INC.
XX Garrett JE, Simin RT, Busby JG, Stormann TM;
XX WPI; 1999-610994/52.
DR P-PSDB; AAY28840.
XX
PT Novel nucleic acids, used to screen for specific modulators, e.g. for
PT treating spasticity or Alzheimer's disease.
XX
PS Disclosure; Fig 3A-3D; 78pp; English.
XX
CC The present sequence is the human calcium receptor nucleic acid sequence.
CC This is used in assaying for compounds modulating GABAB activity. GABAB
CC receptors are metabotropic receptors that modulate synaptic transmission
CC in brain. Portions of calcium receptor are swapped with portions of
CC GABABR2 to create protein fusions. The agonist extracellular binding
CC domain is swapped for measurement of antagonist activity. Intracellular
CC domains are swapped for measuring intracellular effects caused by the
CC different receptor.
XX
SQ Sequence 3234 BP; 739 A; 931 C; 829 G; 735 T; 0 other;

alignment_scores:
  Quality: 161.00 Length: 676
  Ratio: 0.583 Gaps: 27
  Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block:
US-09-775-181-4 x AA90922 ..
Align seg 1/1 to: AA90922 from: 1 to: 3234

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 TCAGACACATTCCTCTACGATTGCTGGTGGGAGCAACTGCTCAGG 443
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 CGTCTCCAGCGCAGTGGCAAACTGCTGGGGCTCTCTACATT..... 486
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
487 .....CCCGAGTCAAGTTATGCTCTCTCCAGC.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuGl 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

568 uillePheAsnMetCysLeuIle 575

|||||

2328 GATCGGCTACACCTGCTGCTG 2349

seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ31049

seq_documentation_block:

ID AAZ31049 standard; cDNA; 3234 BP.

XX AAZ31049;

XX 07-JAN-2000 (first entry)

XX Human Car coding sequence.

XX G-protein fusion receptor; Car; calcium receptor; GluR; head injury;
KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;
KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;
KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;
KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;
cognitive disorder; ss.

XX Homo sapiens.

XX WO9951641-A1.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSF-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;
PI Simin RT;

XX WPI; 1999-610995/52.

XX P-PSDB; AAY49105, AAY49110, AAY49115.

XX New G-protein fusion receptors and chimeras containing domains from
PT different receptors, used to screen for modulators, potentially useful
PT e.g. for treating or preventing stroke or Alzheimer's disease

XX Disclosure: Fig 5; 255pp; English.

XX The invention relates to G-protein fusion receptors (I) comprising:
CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and
CC intracellular (ICD) domains, each chosen independently from a Car
CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR
CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to
CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the
CC linker. (I), and recombinant chimeric receptors (CR) without the GP
CC component, are used to assess function of the various domains and to
CC identify compounds (e.g. allosteric modulators or antagonists) that act
CC on these domains. The modulators are potentially useful for treating or
CC preventing diseases associated with the receptors, e.g. stroke, head or
CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,
CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive
CC disorders and depression. Nucleic acid (II) that encodes (I) is used:
CC (1) for recombinant production of corresponding proteins; and (2) to
CC produce cells used in screening for modulators. Use of Car and mGluR
CC domains allows presentation of GABABR domains, to a binding agent, in a
CC form more like the natural domain structure compared with use of
CC incomplete receptors, lacking one or more domains. By shuffling different
CC domains, agents can be identified that affect particular domains of a
CC receptor.

XX Sequence 3234 BP; 739 A; 931 C; 829 G; 735 T; 0 other:

alignment_scores:

Quality: 161.00

Length: 676

Ratio: 0.583 Gaps: 27

Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block:

US-09-775-181-4 x AAZ31049 ..

Align seg 1/1 to: AAZ31049 from: 1 to: 3234

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
:::|||||:::|||||
394 TCAGACACATTCCTCTACGATTGCTGGTGGGACCACTGGCTCAGG 443
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
: : : : :
444 CGTCTCCACGGCAGTGGCAATCTGCTGGGCTCTCTACATT..... 486
108 euProGlyLysTyrProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
||| : : : : :
487CCCCAGGTCAATTATGCTCTCCCTCCAGC.....AGA 516
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
||| : : : : :
517 CTCCTCAGCAACAAGATCAATTCAAGTCTTCTCCGAACC...ATCCC 563
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTyrG 158
: : : : :
564 CAATGATGAGCAGCAGGCCACTGCCATGGCAGACATCATCGAGTATTTC 613
158 lnAlaLeuValTrpSerLeuLeuGluGluProSerIleSer..... 172
: : : : :
614 GCTGGAACTGGTGGGCACAAATTGCAGCTGATGACGACTATGGGCGGCG 663
173ArgAlaAlaIleThrPheSer 180
||| : : : : :
664 GCGATTGAGAAATCCGAGAGGAGCTGAGAAAGGATATCTGCATCGA 713
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
|| : : : : :
714 CTTCACTGAACATCTCTCCAGTACTCTGATGAGGAAGAGATCCAGCATG 763
196ArgGluGluSer.ArgIleLeuLeuG 204
||| : : : : :
764 TGTAGAGGTGATTCAAAATCCACGGCCCAAGTCACTCGTGGTTTCTCC 813
204 lnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
||| : : : : :
814 AGTGGCCCATCTTGGCCCTCATCAAGGAGATGTCGCGCGCAA... 860
221 GluTrpPheHisGlyLeuArgArgLysTyrArgProHisLeuHisArgAr 237
: : : : :
861TATCAGCGGCAAGATCTGGCTGGCGCAGGCGCTGGCGCAGCT 904
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
||| : : : : :
905 CTCCTCTGATCGGCATGCTCTAGTACTTCCACGTGTTGGCGGCACCAT 954
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
||| : : : : :
955 GGATTGCTCTGAAGC.....TGGGCA 977
267 ProProTyrLeuGlu.....Cy 272
||| : : : : :
978 GATCCCGAGGCTCCGGGAATTCCTGAAGAAGGTCATCCCGAAGTCTG 1027
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
||| : : : : :
1028 TCCACAATGTTTGGCAAGGAGTCTTTGGGAGAAACATTAACTGCCAC 1077
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
||| : : : : :
1078 CTCCAAGAAGGTCCAAAGGACCTTTACCTGTGGACACCTTCTCGAGAG 1127
295AsnLeuValProGluPheArg 302

```
1128 TCACGAAGAAAGTGGCGAGCAGTTTACCAACAGCTCGACAGCCTTCCGAC 1177
      |||
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1178 CCTCTGTACAGGGGATGAGAACATCAGACAGTGTGAGAGCCCTTACATA 1227
313 ..... 313
1228 GATTACAGCATTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1277
314 ..... 318
1278 CATTCGCCACGCCTTGCAGATATATATACCTGCTTACTCGGAGAGGC 1327
319 ..... 335
1328 TCTTCAACAATGGCTCTGTGCGAGACATCAAGAAAGTTGAGGCGTGGCAG 1377
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1378 GTCCTGAAGCACCTACGGCATCTAAACATTTACAAACAATATGGGGAGCA 1427
336 ..... 347
1428 GGTGACCTTTGATGAGTGTGGTACCTGGTGGGAACATTCCATCATCA 1477
348 .....Gly 348
1478 ACTGGCACCTCTCCCGACAGGATGCTCCATCGTGTTTAAGGAAGTCGGG 1527
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
1528 TATTACAAAGCTATATGCCAAGAGGAGAAAGACTCTTCAACAGGAGA 1577
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
1578 GAAATCTGTGGAGTGGGTCTCCAGGAGGTGCGCTTCCCAACTGCA 1627
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
1628 GCCGAGACTGCTGCGAGGACCAAGAGGATCATTTGAGGGGGAGCCC 1677
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
1678 ACCTGCTGCTTGTAGTGTGTGGAGTGTCTGATGGGAGTATAGTGATGA 1727
399 .....CysProPheCysAlaAsp..... 405
1728 GACAGATGCCAGTGCCTGTAAACAAGTCCAGATGACTTCTGGTCCAATG 1777
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
1778 AGAACACACCTCTCGCATTCGCAAGAGATCGAGTTCTGTCTGGGAGC 1827
418 .....LeuAlaIleSerPheGlnGlyLeuCysMetLeuLe 430
1828 GAGCCCTTTGGGATCGCACTACCTCTTTCGCGTGGCATTTCCT 1877
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
1878 GACAGCCTTGTGCTGGGTGTGTTTATCAAGTTCGCGAACACACCACTTG 1927
447 LeArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
1928 TCAAGCCACCAACAGCAGCTCTCTACCTCTCTCTCTCTCTCTCTCTCT 1977
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
1978 TGCTGCTTCTCCAGCTCCCTGTCTCTTCATCGGGGAGCCCCAGGACTGGAC 2027
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
2078 GTGCGCGCTGCCCGCAGCGCGCCTTTTGGCATCAGCTTCGTGCTCTCATCT 2077
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
2078 CATGCATCTGTGTGAAAACCAACCGGTGCTCTCTGTGTGTGTGTGAGGCCAAG 2127
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
2128 ATCCCCACACAGCTTCCACCGCAAGTGGTGGGGCTCAACCTCAGTTCT 2177
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2178 GCTGTTTTCTCTCTGCACCTTCATGCAGATGTCACTGTGTGATCTGGC 2227
540 heLeuIleGlyThrPheSerVal...CysGlnAsnLeuGluLysGln 555
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
2228 TCTACACGCGCGCCCTCAAGCTACCGCAACGACGAGCTGGAGGATGAG 2277
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2278 ATCATCTTCATCAGCTGCCACGAGGCTCCCTCATGCGCCCTGGCTTCT 2327
568 uIlePheAsnMetCysLeuIle 575
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2328 GATCGGCTACACCTCCCTGCTG 2349
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ51398
seq_documentation_block:
ID_ AAZ51398 standard; cDNA; 3234 BP.
XX AC AAZ51398;
XX DT 20-JUN-2000 (first entry)
XX DE Human wild type calcium sensing receptor, CaSR encoding cDNA.
XX KW Calcium Sensing Receptor; CaSR; human; calcium level; modulator;
XX KW hypotensive; osteopathic; receptor activity; hyperparathyroidism;
XX KW osteoporosis; Paget's disease; treatment; hypercalcaemia malignancy;
XX KW hypertension; gene therapy; ss.
XX OS Homo sapiens.
XX KE 1..3234 Location/Qualifiers
XX FT CDS /*tag= a
XX FT /product= "Human wild type calcium sensing receptor"
XX FT /note= "Does not include stop codon"
XX FT /partial
XX PN W0200006601-A1.
XX PD 10-FEB-2000.
XX PF 28-JUL-1999; 99WO-US17116.
XX PR 30-JUL-1998; 98US-0094702.
XX PA (AVET ) AVENTIS PHARM PROD INC.
XX PI Yu KT, Labaudiniere RF, Thrower LW;
XX DR WPI; 2000-195263/17.
XX DR P-PSDB; AAY70325.
XX PT Nucleic acids encoding isoforms of human calcium sensing receptor for
XX PT treating, e.g. hyperparathyroidism or osteoporosis -
XX PS Claim 1; Page 70-76; 81pp; English.
XX CC The present sequence is the cDNA encoding the wild type human calcium
XX CC sensing receptor CaSR, expressed in the kidney. There are multiple
```

CC alternatively spliced isoforms of CaSR, that arise from partial deletion
 CC of nucleotides from this wild type sequence. The CaSR isoforms have
 CC hypotensive and osteopathic activity. The isoforms can be used to
 CC identify agonists and antagonists that modulate the receptor activity
 CC and calcium levels. These modulators are useful for treating
 CC hyperparathyroidism or osteoporosis, Paget's disease, hypercalcaemia
 CC malignancy or hypertension. The DNA sequence is also useful for altering
 CC the CaSR activity and in gene therapy.

XX
 SQ Sequence 3234 BP; 740 A; 930 C; 828 G; 736 T; 0 other;

alignment_scores:

Quality: 161.00 Length: 676
 Ratio: 0.583 Gaps: 27
 Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block:

US-09-775-181-4 x AAZ51398 ..

ign seg 1/1 to: AAZ51398 from: 1 to: 3234

```

75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   ::::::::::::::::::::
394 TCAGAGCACATTCCTCTACGATGCTGTGCTGGAGCAACTGGCTCAGG 443
   ::::::::::::::::::::

91 rHisGlnLeuLeuArgAlaAsnCyssSerGlyArgTyrGluLeuAlaGlyL 108
   ::::::::::::::::::::
444 CGTCTCCACGGCAGTGGCAAAATCTGCTGGGGCTCTCTACATT..... 486
   ::::::::::::::::::::

108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   ::::::::::::::::::::
487 .....CCCCAGTGCAGTATGCTCTCCCTCCAGC.....AGA 516
   ::::::::::::::::::::

125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   ::::::::::::::::::::
517 CTCTCAGCAACAGAAATCAATTCAGTCTTTCTCTCCGAACC...APCC 563
   ::::::::::::::::::::

141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
   ::::::::::::::::::::
564 CAATGATGAGCAGCAGGCCACTGCCATGGCAGACATCATCGAGTATTCC 613
   ::::::::::::::::::::

158 ln.AlalaValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   ::::::::::::::::::::
614 GCTGGAACTGGTGGGCACAATTCAGCTGATGACGACTATGGGGCGCCG 663
   ::::::::::::::::::::

173 .....ArgAlaAlaIleThrPheSer 180
   ::::::::::::::::::::
664 GGGATTGAGAAATCCGAGGAGAGCTGAGGAAAGGATATCTGCATCGA 713
   ::::::::::::::::::::

180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   ::::::::::::::::::::
714 CTTGAGTGAATCATCTCCAGTACTCTGATGAGGAGAGATCCAGCATG 763
   ::::::::::::::::::::

196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   ::::::::::::::::::::
764 TGGTAGAGGTGATTCAAAATTCACGGCCAAAGTCACTCGTGGTTTCTCC 813
   ::::::::::::::::::::

204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   ::::::::::::::::::::
814 AGTGGCCCATCTTGGCCCTTCATCAAGGAGATGTCGGGGCGCAA... 860
   ::::::::::::::::::::

221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   ::::::::::::::::::::
861 .....TATCAGGGCAAGATCTGGCTGGCGCAGCGAGGCTGGGGCAGCT 904
   ::::::::::::::::::::

237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerTr 250
   ::::::::::::::::::::
905 CTTCCCTGATCGGCATTCCTCAGTACTTCCACGTGTTGGCGGCACCAT 954
   ::::::::::::::::::::

250 rParArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   ::::::::::::::::::::
955 GGATTCGCTCTGAAGC.....TGGGCA 977
   ::::::::::::::::::::

```

```

267 ProProTyrLeuGlu.....Cy 272
   ::::::::::::::::::::
978 GATCCAGGCTTCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1027
   ::::::::::::::::::::

272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   ::::::::::::::::::::
1028 TCCACAATGTTTGGCAAGAGAGTTTGGGAAGAAACATTTAACTGCCAC 1077
   ::::::::::::::::::::

285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   ::::::::::::::::::::
1078 CTCAGAAGAGTGCAAAAGGACCTTTACCTTGGACACCTTTCTGAGAGG 1127
   ::::::::::::::::::::

295 .....AsnLeuValProGluPheArg 302
   ::::::::::::::::::::
1128 TCAGGAAGAAGTGGGACAGAGTTTAGCAACAGCTCGACAGCCTTCGAC 1177
   ::::::::::::::::::::

302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   ::::::::::::::::::::
1178 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1227
   ::::::::::::::::::::

313 ..... 313
   ::::::::::::::::::::

1228 GATTACAGCATTTACGGATATCTCAATGTGTACTTAGCAGTCTACTC 1277
   ::::::::::::::::::::

314 .....AspIleAspGlnCys..... 318
   ::::::::::::::::::::
1278 CATTCGCCAGCCTTCAAGATATATATACCTGCTTACCTGGGAGAGGC 1327
   ::::::::::::::::::::

319 .....SerSerAspGlyTrpPhe 324
   ::::::::::::::::::::
1328 TCTTCAACATGGCTCTCTGTGCACATCAAGAAAGTTGAGCGCTGCCAG 1377
   ::::::::::::::::::::

325 SerGlyThrHisLysCysHisLeuAsnSer..... 335
   ::::::::::::::::::::
1378 GTCTGAAGCCTACCGCATCTAACTTTTACAAACAATATGGGGGAGCA 1427
   ::::::::::::::::::::

336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
   ::::::::::::::::::::
1428 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGGAACATTTCCATCATCA 1477
   ::::::::::::::::::::

348 .....Gly 348
   ::::::::::::::::::::
1478 ACTGGCACCTCTCCAGAGAGTGGCTCCATCGTGTAAAGGAAGTCGGG 1527
   ::::::::::::::::::::

349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
   ::::::::::::::::::::
1528 TATTACAAGCTCTATGCCAAGAGGAGGAAAGACTCTTCATCAACGAGGA 1577
   ::::::::::::::::::::

362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
   ::::::::::::::::::::
1578 GAAATCCTGTGGAGTGGTTCCTCCAGGGAGGTGCCCTTCTCCAACTGCA 1627
   ::::::::::::::::::::

378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   ::::::::::::::::::::
1628 GCGGAGACTGCTGGCAGGAGCAGGAAAGGATCATTCAGGGGAGGCC 1677
   ::::::::::::::::::::

389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   ::::::::::::::::::::
1678 ACCTGCTGCTTTGAGTGTGTGAGTGTCTGTATGGGGAGTATAGTATGA 1727
   ::::::::::::::::::::

399 .....CysProPheCysAlaAspAsp..... 405
   ::::::::::::::::::::
1728 GACAGATGCCAGTGCCTGTAAACAAGTGCCAGATGACTTCTGTGTCCAATG 1777
   ::::::::::::::::::::

406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
   ::::::::::::::::::::
1778 AGAACACACCTCTCTGCATTTGCCAAGAGATCGAGTTTCTGCTGGAGC 1827
   ::::::::::::::::::::

418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
   ::::::::::::::::::::
1828 GAGCCCTTTGGGATCCACTACCCCTCTTTTGGCGTCTGGGCATTTTCCT 1877
   ::::::::::::::::::::

```



```
|||||
2700 GATCGCTACACCTGCCTCTG 2721
seq_name: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:AAV26964
seq_documentation_block:
ID AAV26964 standard; cDNA to mRNA; 3809 BP.
XX AC AAV26964;
XX DT 01-SEP-1998 (first entry)
XX DE Human parathyroid calcium receptor 4.0 gene 4Kb fragment.
XX KW ss; calcium ion concentration; parathyroid hormone; homeostasis;
XX KW kidney; calcium receptor; detection.
XX OS Homo sapiens.
XX FT
XX Key Location/Qualifiers
XX CDS 373..3809
XX /*tag= a
XX /product= "pHuPcar 4.0 4Kb fragment"
XX US5763569-A.
XX 09-JUN-1998.
XX 07-JUN-1995; 95US-0484565.
XX 07-JUN-1995; 95US-0484565.
XX 23-AUG-1991; 91US-0749451.
XX 11-FEB-1992; 92US-0834044.
XX 21-AUG-1992; 92US-0934161.
XX 12-FEB-1993; 93US-0017127.
XX 23-FEB-1993; 93US-0009389.
XX 22-OCT-1993; 93US-0141248.
XX 19-AUG-1994; 94US-0292827.
XX 21-OCT-1994; 94WO-US12117.
XX 08-DEC-1994; 94US-03353784.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX (NPSP-) NPS PHARM INC.
XX PI Brown EM, Garrett JE, Hebert SC;
XX WPI; 1998-347412/30.
XX P-PSDB; AAW54846.
XX Calcium receptor poly:peptide(s) - useful for drug screening or
XX antibody production
XX Example 27; Fig 49; 174pp; English.
XX The Human parathyroid calcium receptor gene encodes a 1078 amino acid
XX protein. The tissue from which the receptor and receptors from bovine
XX parathyroid and rat kidney are derived, respond to changes, and control
XX changes, in calcium ion concentration, e.g. parathyroid hormone regulates
XX Ca2+ homeostasis in blood and extracellular fluid, and kidney function
XX alters through changes in Ca2+ levels in juxtaglomerular and proximal
XX tubule cells in the kidney. The purified receptors (produced
XX recombinantly) can be used to screen for compounds that modulate calcium
XX receptor activity, especially those that can be used to treat diseases
XX associated with the receptors in these tissues. They can also be used
XX to raise antibodies for use in detection assays.
XX Sequence 3809 BP; 910 A; 1071 C; 979 G; 849 T; 0 other;
XX
XX alignment_scores:
XX Quality: 161.00
XX Ratio: 0.583
XX Length: 676
XX Gaps: 27
```

```
Percent Similarity: 40.828 Percent Identity: 20.118
alignment_block:
US-09-775-181-4 x AAV26964
Align seg 1/1 to: AAV26964 from: 1 to: 3809
75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
766 TCAGAGCACAATTCCTCTACGATTGCTGGTGGGAGCAACATGGCTCAGG 815
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
816 CGTCTCCAGCGGAGTGGCAATCTGCTGGGCTCTCTACATT
858 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
859 CCCCAGGTCAAGTATGCTCTCCAGC
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuGl 141
889 CTCCTCAGCAACAAGATCAATTCAGTCTTCTCCGAACC...ATCCC 935
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTyrTrpG 158
936 CAATGATGAGCAGCCAGCCACTGCCATGCCACACATCATCGAGTATTTCC 985
158 InAlaLeuValTrpSerLeuLeuGlyGluProSerIleSer
986 GCTGGAACCTGGTGGGCACAATTCAGCTGATGACGACTATGGCGGCGG 1035
173 ArgAlaAlaIleThrPheSerT 180
1036 GGGATTGAGAAATCCGAGAGGAAGCTGAGGAAAGGGATATCGCATCGA 1085
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr
1086 CTTCACTGAATCTCTCCAGTCTGATGAGGAGATTCGCCGCGCAA... 1232
196 ArgGluGluSerArgIleLeuLeuG 204
1136 TGGTAGAGGTGATTCAAATTCACGGCCAGGAGGCTGCTGGTTTCTCC 1185
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
1186 AGTGGCCAGATCTTGAGCCCTCATCAGAGGAGATTGCCGCGCAA... 1232
221 GluTyrPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237
1233 TATCAGGGCAAGATCTGGCTGCCAGGAGGCTGGCCAGCT 1276
237 gGlyProAsnGlnGly
1277 CCTCCCTGATGCCCATCCCTCAGTACTTCCACGTGGTGGCGGCACCAT 1326
250 rParArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
1327 GGATTCGCTCTGAAGGC
267 ProProTyrLeuGlu
1350 GATCCACGCTTCGGGAATTCCTGAAGAGTCCATCCAGGAAGTCTG 1399
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr
1400 TCCACAATGGTTTGCAGAGGAGTTTGGGAAGAAACATTTAACTGCCAC 1449
LeuSerSerAlaIleTyrGlyLeuGlnPro
1450 CTCCAAGAGGTGCAAAAGGACCTTTTACCTGTGGACACCTTTCTGAGAGG 1499
295 AsnLeuValProGluPheArgG 302
```

```
1500 TCACGAAGAAGTGGCGACAGGTTTACGACAGCTCGACAGCCCTCCGAC 1549
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1550 CCCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGACCCCTTACATA 1599
313 ..... 313
1600 GATTACACGCATTTACGGATATCCTACAATGTGTACTTAGCAGTCTACTC 1649
314 .....AspIleAspGlnCys..... 318
1650 CATTGCCACCCCTTGAAGATATATATACCTGCTTACCTGGGAGAGGGC 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTTACCAATGGCTCTGTGTGCACATCAAGAAAGTTGAGGCGTGGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1750 GTCTGAAGCACCTTACGGCATCTAACTTTACAACAATATGCGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGAATATATCCATCAICA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCCGACAGATGGCTCCATCGTTTAAAGGAAGTCGG 1899
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGly 362
1900 TATTACAACGCTATATGCCAAGAGGAGAAAGACTCTTATCAACAGGAGA 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
1950 GAATCTCTGTGGAGTGGTTCACAGGAGGTCCTTCTCCAACATGCA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCCTGGCAGGACCAAGGAGGATCATTGAGGGGAGGCC 2049
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGTTGAGTGTGTGGAGTGTCTCATGGGAGTATAGTATGA 2099
399 .....CysProPheCysAlaAspAsp..... 405
2100 GACAGATGCCAGTGCCTGTACAGATGCCAGATGACTTCTGGTCCATG 2149
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACCTCTCTGCATTTGCCAAGGAGATCGAGTTTCTGTGCGGAGC 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATGCGACATCACCCCTCTTTTGGCGTGTGGGCAATTCCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTGTGGTGTGTATTACAGTTCCGCAACACACCATG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACCGAGACTCTCTACCTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGTGCTTCTTCCAGCTCCCTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 GTCCGCGCTCGCCAGCGCCCTTTGGCATCATCAGCTGTGCTGTGCTGTCT 2449
```

```
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2450 CATGCATCTGGTGAACAAACCAACCGTCTCTCTCTCTCTCTCTCTCTCT 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2500 ATCCCCACCACTTCCACCGCAAGTGGTGGGGCTCAACCTGCAGTTCTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrp 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2550 GCTGGTTTCTCTCTGACCTTCATGACAGATTGTCATCTGTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2600 TCTACACCGCGCCCTCAAGCTACCGCAACGAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2650 ATCATCTTCATCAGTGCACGAGGGCTCCCTCATGGCCCTGGGCTTCTCT 2699
568 uIlePheAsnMetCysLeuIle 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2700 GATCGGCTACCTGCTGCTG 2721

seq_name: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:AAT95859
seq_documentation_block:
ID AAT95859 standard; cDNA to mRNA; 3809 BP.
XX
AC AAT95859;
XX
DT 08-MAY-1998 (first entry)
XX
DE Human parathyroid cell calcium receptor 4.0 (HuPcAr 4.0) cDNA.
XX
KW Human parathyroid cell calcium receptor 4.0; HuPcAr 4.0;
KW calcium homeostasis; hyperparathyroidism; osteoporosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 373..3609
FT FT /*tag= a
FT FT /product= parathyroid_cell_calcium_receptor_4.0
XX
PN US5688938-A.
XX
PD 18-NOV-1997.
XX
PF 07-JUN-1995; 95US-0485588.
XX
PR 07-JUN-1995; 95US-0485588.
PR 23-AUG-1991; 91US-0749451.
PR 11-FEB-1992; 92US-0834044.
PR 21-AUG-1992; 92US-0934161.
PR 12-FEB-1993; 93US-0017127.
PR 23-FEB-1993; 93US-0009389.
PR 22-OCT-1993; 93US-0141248.
PR 19-AUG-1994; 94US-0292827.
PR 21-OCT-1994; 94WO-US12117.
PR 08-DEC-1994; 94US-0353784.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PA (NPSF-) NPS PHARM INC.
XX
PI Brown EM, Fuller FH, Garrett JE, Hebert SC;
XX
WPI; 1998-008040/01.
DR P-PSDB; AAW38274.
XX
PT DNA encoding calcium receptor polypeptide(s) - useful for
therapeutic purposes, e.g. hyperparathyroidism and osteoporosis
```

xx Claim 15; Columns 125-134; 174pp; English.
 xx The present sequence encodes human parathyroid cell calcium
 CC receptor 4.0 (hupCar 4.0).
 CC The specification includes details of molecules that can modulate
 CC one or more inorganic ion receptor activities, and antibodies and
 CC antibody fragments targetted to inorganic ion receptor proteins. The
 CC proteins, nucleic acids and antibodies may be used to treat
 CC disorders by modulating one or more inorganic ion receptor
 CC activities, preferably disorders of calcium homeostasis, e.g.
 CC hyperparathyroidism and osteoporosis.
 xx
 SQ Sequence 3809 BP; 910 A; 1071 C; 979 G; 849 T; 0 other;

alignment_scores:
 Quality: 161.00 Length: 676
 Ratio: 0.583 Gaps: 27
 Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block:
 US-09-775-181-4 x AAT95859 ..
 Align seg 1/1 to: AAT95859 from: 1 to: 3809

```

75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 TCAGAGCATTCCCTCAGATTGCTGTGGTGGAGCAATGGCTCAGG 815
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
816 CGTCTCCACGGCAGTGGCAAAATCTGCTGGGCTCTTCTACATT..... 858
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
859 .....CCCGAGTTCAGTTATGCTCTCTCCAGC.....AGA 888
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
889 CTCCTCAGCAACAGAAATCAATTCAGTCTTTCCGCCGAACC...ATCCC 935
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
936 CAATGATGAGCACCAGGCCACTGCCATGGCAGACATCATCGAGTATTCC 985
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
158 InAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
986 GCTGGAACGTGGTGGGCACAAATTCGAGCTGATGACGACTATGGGCGCGG 1035
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
173 .....ArgAlaAlaIleThrPheSerT 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 GGGATTGAGAAATCCGAGAGGAGCTGAGGAAGGATATCTGTCATCGA 1085
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 CTTAGTGAACATCTCCAGTACTCTGTATGAGGAGAGATCCAGCATG 1135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....ArgGluGluSerArgIleLeuLeuG 204
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 TGGTAGAGGTGATTCAAATTCACGGCCCAAGTATCTGTGGTTTCTHCC 1185
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 AGTGCCCCAGATCTTGAGCCCTCATCAAGGAGATTGTCGCGCGCAA... 1232
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1233 .....TATCAGGGCAAGATCTGCTGGCCAGCGAGGCTGGGCCAGCT 1276
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1277 CCTCCCTGATGCCATGCTCAGTACTTCCACGTTGTTGGCGGCCACCAAT 1326

```

```

250 rpArgArgLysAspGlyLeuGlyAspLysSerHisPheLysTrpSer 266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1327 GGATTGCTGCTGAGGC.....TGGGCA 1349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ProProTyrLeuGlu.....Cys 272
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1350 GATCCCAAGCTTCGGGAATTCCTCAAGAAGCTCCATCCAGGAAGTCTG 1399
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1400 TCCACAATGGTTTGGCAAGGAGTTTGGGAAGAACAATTTAACTGCCAC 1449
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTGAGAG 1499
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....AsnLeuValProGluPheArgG 302
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1500 TCACGAAGAAAGTGGCGACAGGTTTAGCAACAGCTCGACAGCCTTCCGAC 1549
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1550 CCCTCTGTACAGGGATGAGAACATCAGCAGTGTGCGAGACCCCTTACATA 1599
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
313 .....SerSerAspGlyTrpPhe 324
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1700 TCTTCAACAATGGCTCTGTCGACAGATCAACAAGTTGAGCGCTGGCAG 1749
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1750 GTCTTGAAGCACCCTACGGCATCTAAACTTTACAAACAATATATGGGAGCA 1799
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1800 GGTGACCTTGTGATGAGTGTGTGACCTGTGGGGAACATTCCATCATCA 1849
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
348 .....Gly 348
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1850 ACTGGCACCCTCCCGACAGGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1899
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG 362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1949 TATTACAAGCTATATGCCAAGAAGGAGAAAGACTCTTTCATCAACGAGGA 1949
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1950 GAAATCTGCTGGAGTGGGTCTCCAGGAGGTGCCCTTCTCCAACTGCA 1999
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2000 GCCGAGACTGCTGGCAGGACCAGGAAGGATCATTTGAGGGGAGGCC 2049
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
389 .....AlaTyrValCysLeuProCysArgGly..... 398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2050 ACCTGCTGCTTGTAGTGTGTGGAGTGTCTGTATGGGAGTATAGTGATGA 2099
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
399 .....CysProPheCysAlaAspAsp..... 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2100 GACAGATCCAGTGCCTGTAAACAGTGCACAGTACTTCTGGTCCAATG 2149
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2150 AGAACCCACACCTCTCTGCATTCGCAAGGAGATCGAGTTTCTGCTGGAGG 2199

```



```
173 .....ArgAlaAlaIleThrPheSerT 180
1036 GGGATTGAGAAATTCGGAGGAGAGCTGAGAAAGGATATATCGTCATGA 1085
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
1086 CTTCACTGAACATCATCTCCAGTACTCTCATGAGGAGAGATCCAGCATG 1135
196 .....ArgGluGluSer.ArgIleLeuLeuLeuG 204
1136 TGGTAGAGGTGATTCAAAATTCACGGGCAAAAGTCATCGTGGTTTCTCC 1185
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
1186 AGTGGCCCAAGATCTTGAGGCTCCATCAAGAGGATGTGCCGCGCAA... 1232
221 GluTrpPheHisGlyLeuArgLysTrpArgProHisLeuHisArgAr 237
1233 .....TATCAGGGCAGATCTGGCTGCCAGCGAGGCTGGGCCAGCT 1276
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
1277 CCTCCCTGATGCCATGCCCTCAGTACTTCCACGTGGTGGCGCACCAT 1326
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
1327 GATTGCTGCTGAAGGC.....TGGGCA 1349
267 ProProTyLeuGlu.....Cy 272
1350 GATCCAGGCTTCGGGAATTCCTGAAGAGGTCCATCCAGCAAGATCTG 1399
272 s.GluAsnGlySerTyrlsProGlyTrpLeuValThr..... 284
1400 TCCACAATGGTTTGCAGAGGATTTGGGAAGAAACATTTAACTGCCAC 1449
285 LeuSerSerAlaIleTyrcGlyLeuGlnPro..... 294
1450 CTCCAAGAGGTGCAAAAGGACCTTTTACCTGTGGACACCTTTCTGAGAGG 1499
295 .....AsnLeuValProGluPheArgG 302
1500 TCACGAAGAAAGTGGCGAGCTTTAGCAACAGCTCGACAGCCTTCGGAC 1549
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
1550 CCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGCCCTTACATA 1599
313 ..... 313
1600 GATTACAGCATTTACGGATATCTTACATGTGTACTTAGCAGTCTACTC 1649
314 .....AspIleAspGlnCys..... 318
1650 CATTCGCCACGCTTCGAAGATATATATACCTTACCTTACCTGGGAGAGGC 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTCACCAATGGCTCCTGTGCAGACATCAAGAAAGTTGAGCGGTGGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
1750 GTCCTGAAGACCATCAGGCACTTAAACTTTACAAACAATATATGGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GTTGACCTTTGATGAGTGTGGTACCTGGTGGGAACATTTCCATCATCA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCAGAGGATGGCTCCATCGTGTGTTAAGGAAGTCGG 1899
349 AlaTyrrGluCysIleCysLysAlaGly.....PheTyrrHisProG 362
```

```
1900 TATTACACGCTCTATGCCAAGAGGAGAAAGACTCTTCTCATCAACAGGA 1949
362 yValLeuProValAsnPheArgArgGlyProAspGlnHis.... 377
1950 GAAATCTCTGTGGAGTGGTTCCTCCAGGAGGTGCCCTTCTCCAACATGCA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCTGGCAGGACAGGAAAGGATCATTTAGGGGGAGGCC 2049
389 .....AlaTyrrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGCTTGGAGTGTGGAGTCTCTGATGGGAGATATAGTATGA 2099
399 .....CysProPheCysAlaAsp..... 405
2100 GACAGATGCCAGTGCCTGTAAACAAGTGCCAGATGACTTCTGTGTCCAATG 2149
406 .....SerProCysPheValGlnGluAspLysTyrrLeuArg..... 417
2150 AGAACACACACCTCTCGCATTTGCCAAGAGATCGAGTTTCTGTGTGGAGC 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTACCTCTTTTCCGCTGCTGGGCATTTTCT 2249
430 uAspPheValSerMetLeuValValTyrrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTGCTGGGTGTGTTTATCAAGTTCGCCACACACCCATTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACCGAGAGCTCTCTACTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrrPheProValIleLeuTyrrPheGluProSerThrPheAr 480
2350 TGCTGCTCTCTCCAGCTCCCTGTTCATCGGGGAGGCCCGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrrG 497
2400 GTGCGCCCTGCGCCAGCGCCCTTTGGCATCAGCTTCGTGCTCGCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCTCTGTGTAACCAACCGTGTCTCTCTGTGTGTGGGCAAG 2499
514 ThrAlaGlnArgIleProTyrrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACACAGCTTCCACCGCAAGTGGTGGGGCTCAACCTGCAGTTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
2550 GCTGTTTTTCTCTCGACCTTCATGCAGATTTGTCATCTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
2600 TCTACACGCGCCCTCAAGCTACCGCAACACGAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTTCATCAGCTGCCAGGAGGCTCTCTCATGGCCCTGGCTTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGCTACACCTGCTGCTG 2721
+###
seq_name: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV82485
seq_documentation_block:
ID: AAV82485 standard; cDNA to mRNA; 3809 BP.
XX
AC AAV82485;
```

| | | |
|----|---|---------------------|
| XX | 19-MAR-1999 | (first entry) |
| DT | | |
| XX | | |
| XX | Human parathyroid calcium receptor pHuPCar 4.0 encoding cDNA. | |
| DE | | |
| XX | | |
| KW | Parathyroid calcium receptor; inorganic ion receptor; osteoporosis; | |
| KW | calcium homeostasis; hyperparathyroidism; seizure; stroke; epilepsy; | |
| KW | spinal cord injury; hypoxia-induced nerve cell damage; cardiac arrest; | |
| KW | neonatal distress; neurodegenerative disease; Alzheimer's disease; | |
| KW | Huntington's disease; Parkinson's disease; dementia; muscle tension; | |
| KW | depression; anxiety; ss. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| XX | Key | Location/Qualifiers |
| FF | CDS | 373..3609 |
| FT | | /*tag= a |
| FT | | |
| XX | | |
| XX | US5858684-A. | |
| PN | | |
| XX | 12-JAN-1999. | |
| XX | | |
| XX | 07-JUN-1995; | 95US-0480751. |
| XX | | |
| PR | 07-JUN-1995; | 95US-0480751. |
| PR | 23-AUG-1991; | 91US-0749451. |
| PR | 11-FEB-1992; | 92US-0834044. |
| PR | 21-AUG-1992; | 92US-0934161. |
| PR | 13-FEB-1993; | 93US-0017127. |
| PR | 23-FEB-1993; | 93US-0009389. |
| PR | 22-OCT-1993; | 93US-0141248. |
| PR | 19-AUG-1994; | 94US-0292827. |
| PR | 18-OCT-1994; | 94WO-US12117. |
| PR | 08-DEC-1994; | 94US-0353784. |
| XX | | |
| XX | (BGHM) BRIGHAM & WOMENS HOSPITAL. | |
| PA | (NPSP-) NPS PHARM INC. | |
| PA | | |
| XX | Balandrin MF, Brown EM, Del Mar EG, Garrett JE; | |
| PI | Hebert SC, Nemeth EF, Van Wagenen BC; | |
| PI | | |
| XX | WPI; 1999-119871/10. | |
| DR | P-PSDB; AAW89565. | |
| XX | | |
| XX | Screening for calcium receptor-active compounds - by recombinant | |
| PT | expression of nucleic acid encoding calcium receptor and determining | |
| PT | the effect of compounds on calcium receptor activity | |
| PT | | |
| XX | Claim 1; Fig 49; 176pp; English. | |
| XX | | |
| CC | A method has been developed of screening for a compound able to affect | |
| CC | one or more activities of a calcium receptor (CR) comprises: (A) | |
| CC | contacting a recombinant cell with a test compound, where the | |
| CC | recombinant cell comprises a recombinant nucleic acid expressing the CR, | |
| CC | provided that the cell does not have functional CR expression from | |
| CC | endogenous nucleic acid; (B) determining the ability of the test | |
| CC | compound to affect one or more activities of the calcium receptor; and | |
| CC | (C) comparing the ability with the ability of the test compound to | |
| CC | affect the one or more CR activities in a cell not comprising the | |
| CC | recombinant nucleic acid. The present sequence encodes human | |
| CC | parathyroid CR, designated a pHuPCar 4.0. The nucleic acid sequence of | |
| CC | pHuPCar 4.0 can be used as part of the recombinant nucleic acid in the | |
| CC | method described above. The compounds identified can be used to treat | |
| CC | diseases or disorders characterised by abnormal calcium homeostasis, e.g. | |
| CC | hyperparathyroidism, osteoporosis and other bone and mineral-related | |
| CC | disorders. They can also be used for the treatment of diseases and | |
| CC | disorders associated with disrupted Ca2+ responses, e.g. seizures, | |
| CC | stroke, spinal cord injury, hypoxia-induced nerve cell damage such as in | |
| CC | cardiac arrest or neonatal distress, epilepsy, neurodegenerative | |
| CC | diseases such as Alzheimer's disease, Huntington's disease and | |
| CC | Parkinson's disease, dementia, muscle tension, depression, and anxiety. | |
| XX | | |
| SQ | Sequence 3809 BP; 910 A; 1071 C; 979 G; 849 T; 0 other; | |

```
|||||..... ||| |||
1450 CTCCAAGAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTCGAGAGG 1499
295 ..... AsnLeuValProGluPheArgG 302
1500 TCACGAAGAAAGTGGCAGACAGTTTATAGCAACAGCTCGACAGCTTCCGAC 1549
302 lyValMetLysValAspLeuGlnLysVal..... 313
1550 CCTCTGTACAGGGGATGAGAACATCATCAGCAGTGTGAGACCCCTTACATA 1599
313 ..... 313
1600 GATTACACGATTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1649
314 ..... AspIleAspGlnCys..... 318
1650 CATTCGCCACGCTTGCAGATATATATACCTGCTTACCTGGGAGAGGCG 1699
319 ..... SerSerAspGlyTrpPhe 324
1700 TCTTCACCAATGGCTCTCTGTGCAGACATCAAGAAAGTTGAGGCGTGGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnSer..... 335
1750 GTCGTGAAGACCTACGGCATCTAAACATTTACAAACAATATGGGGGAGCA 1799
336 ..... GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGCTGACCTGGTGGGAACACTTCCATCATCA 1849
348 ..... Gly 348
1850 ACTGCGACCTCTCCCGACAGGATGCTCCATGCTGTTTAAGGAAGTCGGG 1899
349 AlaTyrGluCysIleCysLysAlaGly..... PheTyrHisProGlu 362
1900 TATTACACGCTATGCTCAAGAGGGAAGAAAGACTCTTCATCAACAGGAGA 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHls.... 377
1950 GAAATCTGTGGAGTGGTCTCCAGGAGGTGGCTTCTCCAACTGCA 1999
378 ..... IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTGCTGCGAGGGACAGAGAAAGGATCATTTGAGGGGAGGCC 2049
389 ..... AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGTTTGGAGTGTGGAGTGTCTCTGATGGGGAGTATAGTGATGA 2099
399 ..... CysProPheCysAlaAspAsp..... 405
2100 GACAGATGCCAGTGCCTGTATCAAGTGGCCAGATGACTTCTGTCCTCAATG 2149
406 ..... SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACACCTCTGCTGATGCAAGAGAGATCGAGTTCTGTCGCGGAGC 2199
418 ..... LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTACCTCTTTGCGCGTGGGCAATTTCTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACACCCCTTTGCTGGGTGGTGTATCAAGTTCGCCAACACACCCCATGTG 2299
447 leArgAlaSerGlyLeuIleLeuGluGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACAGAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValValIleLeuTyrPheGluProSerThrPheAr 480
```

```
2350 TGCTGCTTCTCCAGCTCCCTGTTCTTTCATCGGGGAGCCCCAGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuGlyPheAlaThrValTyrG 497
2400 GTGCGCCTGCGCCAGCGGCTTTGGCATCAGCTTCGTGCTCTCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCTGTTGTAACCAACACCGCTGCTCTGCTGTTTGGAGGCCAAG 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACACGCTTCCACCGCAAGTGTGGGGCTCAACCTGCAGTTCTCT 2549
530 uAlaVal..... IleLeuLeuValValPheTrpP 540
2550 GCTGTTTCTCTGCACCTTTCATGCAGATTGTCATCTGTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal... CysGlnAsnLeuGluLysGln 555
2600 TCTACACCGCGCCCTCAAGCTACCGCAACCCAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle..... GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGCTGCACGAGGCTCCCTCATGCGCCTGGCTTCTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGCCTGCTG 2721
seq_name: /SIDSL/gcgdata/geneseq/NA2000.DAT:AAZ89298
seq_documentation_block:
ID AAZ89298 standard; cDNA to mRNA; 3809 BP.
XX
AC AAZ89298;
XX
DT 09-JUN-2000 (first entry)
XX
DE Human calcium receptor phuCar4.0 cDNA.
XX
KW Calcium receptor; treatment; calcimimetic; calcilytic; osteopathic;
KW cerebroprotective; cytosolic; neuroprotective; dermatological;
KW tranquilizer; vulnary; antitumor; immunosuppressive; hypotensive;
KW cardant; parathyroid hormone; osteoporosis; calcitonin secretion;
KW hyperparathyroidism; Paget's disease; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 373..3609
FT FT /*tag= a
FT FT /product= "phuCar4.0"
XX
PN US6031003-A.
XX
PD 29-FEB-2000.
XX
PF 07-JUN-1995; 95US-0484719.
XX
PR 23-AUG-1991; 91US-0749451.
PR 17-FEB-1992; 92US-0834044.
PR 21-AUG-1992; 92US-0934161.
PR 12-FEB-1993; 93US-0017127.
PR 23-FEB-1993; 93US-0009389.
PR 22-OCT-1993; 93US-0141248.
PR 19-AUG-1994; 94US-0292827.
PR 21-OCT-1994; 94WO-US12117.
PR 08-DEC-1994; 94US-0353784.
XX
PA (NPSP-) NPS PHARM INC.
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
```



```
348 ..... Gly 348
1850 ACTGGCAGCTCTCCCAAGAGATGCTCCATCGTGTTTAAAGAAAGTCGGG 1899
349 AlatyrgluCysIleCysIleAlaGly ..... PheTyrHisProG1 362
1900 TATTACAGCTCTATGCCAAGAGGGAAGAACTCTTTCATCAGGAGGA 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis .... 377
1950 GAAATCCTGTGGAGTGGGTCTCCAGGAGGTGCCCTTCTCCAACATGCA 1999
378 ..... IleSerGlySerThrLysAspValSerGluGlu ..... 388
2000 GCCGAGACTGCTGCCAGGACCAAGAGGATCATTTGAGGGGAGGCC 2049
389 ..... AlaTyrValCysLeuProCysArgGluGly ..... 398
2050 ACCTGCTGCTTTGAGTGTGTGGAGTGTCTGTATGGGAGTATAGTGAGA 2099
399 ..... CysProPheCysAlaAsp ..... 405
2100 GACAGATGCCAGTCCCTGTACAAGTGCACAGATGACTTCTGTGTCCTCAATG 2149
406 ..... SerProCysPheValGlnGluAspLysTyrLeuArg ..... 417
2150 AGAACACACTCTCTGCTGATCCCAAGGAGATCGAGATTTCTGTGTGGACG 2199
418 ..... LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTCACCTCTTTGCCGTGCTGGGCATTTCTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGCTGTGGTGTGTATCAAGTCCGCAACACACCCATTG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACGAGAGCTTCCTACCTCCCTCTCTCTCCCTGCTC 2349
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGCTGCTTCTCCAGCTCCCTGTTCTTCATCGGGAGCCCAAGGACTGGAC 2399
480 gCysIleLeuLeuArgThrAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 CTGCCGCTGCGCCAGCGGCTTTGGCATCAGCTTCGTCTGCTGCTCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGATCCTCGTGGTGAACCAACCGTGTCTCTGCTGTTTGGAGGCCAAG 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACCACTTCCACCGCAAGTGGTGGGGCTCACTGCTGCTGCTCT 2549
530 uAlaVal ..... IleLeuLeuValValPheThrpp 540
2550 GCTGTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2599
540 heLeuIleGlyTrpThrSerSerVal ..... CysGlnAsnLeuGluLysGln 555
2600 TCTACAGCGCGCCCTCTCAAGTACCGCAACCAAGGAGCTGAGGAGTGA 2649
556 IleSerLeuIle ..... GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGTGCACGAGGGCTCTCTCATGCGCCCTGGGCTTCTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGCCTGCTG 2721
```

seq_name: /SIDSL/gcgdata/geneseq/geneseq/nA2000.DAT:AAZ50617

seq_documentation_block:
ID AAZ50617 standard; cDNA; 2703 BP.

```
XX
AC AAZ50617;
DT 20-JUN-2000 (first entry)
XX
DE Human calcium sensing receptor isoform encoding cDNA.
DE
KW Calcium Sensing Receptor; CaSR; isoform; human; splice variant;
KW hypotensive; osteopathic; receptor activity; calcium level; modulator;
KW treatment; hyperparathyroidism; osteoporosis; Paget's disease;
KW hypercalcaemia malignancy; hypertension; gene therapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Location/Qualifiers
FT 1..2703
FT /tag= a
FT /product= "Human calcium sensing receptor isoform"
FT /note= "Does not include stop codon"
FT /partial
XX
PN WO200006601-A1.
XX
PD 10-FEB-2000.
XX
PF 28-JUL-1999; 99WO-US17116.
XX
PR 30-JUL-1998; 98US-0094702.
XX
PA (AVET ) AVENTIS PHARM PROD INC.
XX
PI Yu KT, Labaudiniere RF, Thrower LW;
XX
WP1: 2000-195263/17.
DR P-PSDB; AAY45001.
XX
PT Nucleic acids encoding isoforms of human calcium sensing receptor for
treating, e.g. hyperparathyroidism or osteoporosis
XX
PS Claim 5; Page -: 81pp; English.
XX
CC The present sequence is the cDNA encoding an isoform of the human
CC calcium sensing receptor CaSR, expressed in the kidney. It is a splice
CC variant of the wild type CaSR, that arise from deletion of nucleotides
CC 1080-1610, from the extracellular domain, that comprise acidic residues.
CC It has hypotensive and osteopathic activity. The CaSR isoforms can be
CC used to identify agonists and antagonists that modulate the receptor
CC activity and calcium levels. These modulators are useful for treating
CC hyperparathyroidism or osteoporosis, Paget's disease, hypercalcaemia
CC malignancy or hypertension. The DNA sequence is also useful for altering
CC the CaSR activity and in gene therapy.
CC Note: This sequence has deletion of nucleotides 1080-1610, according to
CC the sequence shown in the specification, but has been stated as
CC 1075-1608 in the claims. This sequence is not found in the specification
CC but has been constructed from the human CaSR sequence (Seq ID.No.11)
CC found in page 70-76.
XX
SQ Sequence 2703 BP; 595 A; 804 C; 689 G; 615 T; 0 other;
```

alignment_scores:
Quality: 158.50 Length: 567
Ratio: 0.652 Gaps: 23
Percent Similarity: 42.857 Percent Identity: 20.459alignment_block:
US-09-775-181-4 x AAZ50617

Align seg 1/1 to: AAZ50617 from: 1 to: 2703

```
75 AlaGluGluValPrometAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   : : : : :
394 TCAGAGACATTCCTCTAGCAATTCCTGGTGGAGCAACTGGCTCAGG 443
   : : : : :
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   : : : : :
444 CGTCTCCAGCGCATGCAATCTGCTGGGGCTCTTACATT..... 486
   : : : : :
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   : : : : :
487 .....CCCCAGTCAGTTATGCCCTCCTCCAGC.....AGA 516
   : : : : :
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG 141
   : : : : :
517 CTCCTCAGCAACAAGAAATCAATTCAGTCTTTCCTCCGAACC...ATCCC 563
   : : : : :
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
   : : : : :
564 CAATGATGAGCACCAGCCCACTGCCATGGGAGAGCATCATCGAGTATTTC 613
   : : : : :
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   : : : : :
614 GCTGGAACCTGGTGGGCACAATTCGACCTGATGACGACTATGGGCGCCG 663
   : : : : :
173 .....ArgAlaAlaIleThrPheSerT 180
   : : : : :
664 GGGATTGAGAAATTCGAGAGGAGCTGAGGAAAGGATATCGCATCGA 713
   : : : : :
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   : : : : :
714 CTTCACTGAATCATCTCCCACTCTGCTGAGGAGAGATCCAGCATG 763
   : : : : :
196 .....ArgGluGluSerArgIleLeuLeuG 204
   : : : : :
764 TGGTAGAGGTGATTCAAAATTCAGGCCCAAGTATCGTGGTTCCTCC 813
   : : : : :
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   : : : : :
814 AGTGGCCAGATCTTGAGCCCTCATCAAGAGAGATTGCCGGGCGCAA... 860
   : : : : :
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   : : : : :
861 .....TATCAGGGGAGAGATCTGGCTGGCCAGCGGCTGGGCGCAGCT 904
   : : : : :
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   : : : : :
905 CCTCCTGATCGCATGCCTCAGTACTTCCACGTGGTGGCGCACCAT 954
   : : : : :
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   : : : : :
955 GGATTCCGCTGAAGGC.....TGGGCA 977
   : : : : :
267 ProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuVa 283
   : : : : :
978 GATCCAGCGGTT.....CCGGGA..... 995
   : : : : :
283 lThrLeuSerSerAlaIleTyrGlyLeuGlnProAsnLeuValProGluP 300
   : : : : :
996 .....ATTCCCTGAAG 1005
   : : : : :
300 heArgGlyValMetLysValAspIleAsnLeuGlnLysValAspIleAsp 316
   : : : : :
1006 AAGGTCCATCCCGAAGTC.....TGTCGA 1031
   : : : : :
317 GlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysHisLeuA 333
   : : : : :
1032 CAATGGTTTGGCAAGAGATTGGGAGAACATTTAACTGCCACCTCC 1081
   : : : : :
333 snAsnSerGluCysMetProIleLysGlyLeuGlyPheValLeuGlyAla 349
   : : : : :
1082 CCTTCTCCAATGTC.....AGC 1098
   : : : : :
350 TyrGluCysIleCysLysAlaGlyPheTyrHisProGlyValLeuProVa 366
   : : : : :
1099 CGAGAGTCC..... 1107
   : : : : :
366 lAsnAsnPheArgArgGlyProAspGlnHisIleSerGlySerThrL 383
   : : : : :
1108 .....CTGGCAGGGACCAAGGA 1123
   : : : : :
383 ysAspValSerGluGlu.....AlaTyrValCysLeuProCys 395
   : : : : :
1124 AAGGATCAITGAGGGGAGCCCACTGCTGCTTTGAGTGTGTGGAGTGT 1173
   : : : : :
396 ArgGluGly.....CysProPheCy 402
   : : : : :
1174 CCTGATGGGAGTATAGTGTGAGACAGATGCCAGTCCCTGTAACAAGTG 1223
   : : : : :
402 sAlaAspAsp.....SerProCysPheValGlnG 412
   : : : : :
1224 CCCAGATGACTTCTGTGTCGTGGACGAGCCCTTGGGATCGCACTCACCTC 1273
   : : : : :
412 luAspLysTyrLeuArg.....LeuAlaIleIleSer 422
   : : : : :
1274 AGATCGAGTTTCTGTGTCGTGGACGAGCCCTTGGGATCGCACTCACCTC 1323
   : : : : :
423 PheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeuValValTy 439
   : : : : :
1324 TTTGCCCTGCTGGGCAATTTCTTCACAGCCTTTGTGCTGGGTGTGTTAT 1373
   : : : : :
439 rHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuIleLeuLeuG 456
   : : : : :
1374 CAAGTTCGCAACACACACCCATTTGTAAGGCCACCAACCGAGAGCTCTCT 1423
   : : : : :
456 luThrIleLeuPheGlySerLeuLeuLeuTyrPheProValValIleLeu 472
   : : : : :
1424 ACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1473
   : : : : :
473 TyrPheGluProSerThrPheArgCysIleLeuLeuArgTrpAlaArgLe 489
   : : : : :
1474 ATCGGGGAGCCCGAGGACTGGAGTGGCGCTGGCGCCAGCGGCTTTGG 1523
   : : : : :
489 uLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLeuHisArgV 506
   : : : : :
1524 CATCAGCTTCGTCTGTCATCTCATGTCATCTCTGGTGGTGAACCAACCGTG 1573
   : : : : :
506 alLeuLysValPheLeuSerArgThrAlaGlnArgIleProTyrMetThr 522
   : : : : :
1574 TCCTCTCTGGTGTGGAGCCCAAGATCCCCACCAAGTTCACCCGCAAGTGG 1623
   : : : : :
523 GlyGlyArgValMetArgMetLeuAlaVal..... 532
   : : : : :
1624 TGGGGGCTCAACCTGCAGTTCTCTGCTGGTTCCTCTCTGCACTTCATGCA 1673
   : : : : :
533 lIleLeuLeuValValPheTrpPheLeuIleGlyTrpThrSerSerVal. 548
   : : : : :
1674 GATTGTCACTGTGTGATCTGGCTCTACACCGCGCCCTCAAGCTACC 1723
   : : : : :
549 ..CysGlnAsnLeuGluLysGlnIleSerLeuIle.....Gly 560
   : : : : :
1724 GCACCAAGGAGCTGGAGGATGAGATCATCTTCATCAGCTCCACGAGGGC 1773
   : : : : :
561 GlnGlyLysThrSerAspHisLeuIlePheAsnMetCysLeuIle 575
   : : : : :
1774 TCCCTCATGGCCCTGGGCTTCCTGTATCGGCTACACCTGCCTGCTG 1818
   : : : : :
```



```

;
; NAME/KEY: CDS
; LOCATION: 373..3606
; OTHER INFORMATION:
; US-08-485-588-3

```

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 161.00 | Length: 676 |
| Ratio: | 0.583 | Gaps: 27 |
| Percent Similarity: | 40.828 | Percent Identity: 20.118 |

```
alignment_block:
US-09-775-181-4 x US-08-485-588-3 ..
Align seq 1/1 to: US-08-485-588-3 from: 1 to: 3809
```

75 AlaGluGluValProMetAspVaIAlaSerTyrLeuTyrThrGlyAspSe 91
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
766 TCAGACACATTTCCCTCTACGATTTCGTGGTGAGCAACTGGCTCAGG 815

91 rHISGLInLeuLysArgAlaAsnCysSerGlyArgTyrcIuLeuAlaGlyL 108
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
816 CGTCTCCAGCGGAGTGGAATCTCTGGTGGGCTCTTCACATT..... 858
108 euProGlyLystrpproAlaLeuAlaSerAlaHisProSerLeuHisArg 124
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
859CCCCAGGTCAAGTTATGCCTCCTCCAGC.....AGA 888

125 AlaLeuAspThrLeuThrHisAlaThrAsnPhetLeuAsnValMetLeuGl 141
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
889 CTCCTCAGCAACAAGAATAAATTCAAAGTCTTTCCCTCCGAACC...ATCCC 935

141 nSerAsnLysSerArgGluInAsnLeuGlnInAspAspLeuAspTrpTyrg 158
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
936 CAATGATGAGCACACAGGCCACTGCCATCGGACAGACATCATCGAGTATTTC 985

158 ln..AlaLeuValTrpSerLeuLeuGluGlyGluProserlleSer..... 172
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
986 GCTGGAACCTGGGTGGGCACAATTGCAGCTGATGACGACTATGGCGGCGG 1035

173ArgAlaAlaIleThrPheSerT 180
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1036 GGGATTGAGAAATTCGAGAGAAAGCTCGAGAAAGGGAATCTCGATCCA 1085

180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr.... 195
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1086 CTTTCAGTGAACTCATCTCCAGTACTCTGTAGAGGAAGAGATCCAGCATG 1135

196ArgGluGluSer..ArgIleLeuLeuG 204
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1136 TGGTAGAGGTGATTCAAATTCAGGCGCAAGGATCATCGTGGTTTTCTCC 1185

204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1186 AGTGCCCGCATCTTGAGCCCTCATCAAGAGAGATTGTCCGGCCCAA... 1332

221 GluTrpPheHisGlyLeuArgLysTrpArgProHisLeuHisArgAr 237
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1233TATCAGGGCAAGATCTGGCTGGCCGAGCGGCTCGGCCACGT 1276

237 gGlyProasnGlnGly.....ProArgGlyLeuGlyHisSerT 250
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1277 CTCTCCTGATCGCCATGCGCTCAGTACTTCCACGTGTGGTGGCGCACCAT 1326

250 rpArgGlyAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1327 GGATTCTCGTCTGAAGGC.....TGGGCA 1349

267 ProProTyrrLeuGlu.....Cy 272
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1350 GATCCACGGCTTCGGGGAATTCCTGAAGAAGGTCCATCCACGGAAGTCTG 1399

272 s.GluAsnGlySerTyrLeuProGlyTrpLeuValThr..... 284
 1400 TCCACAATGGTTTGGCCAAGAGGATTTTGGGAAGAAACATTTAACATGCGCAC 1449
 285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
 1450 CTCCAAGAAGGTCGAAAGAGCCTTTACCTGTGGACACCTTTCTGAGAGG 1499
 295AsnLeuValProGluPheArg 302
 1500 TCACGAAGAAGTGGCGACAGGTTTACACACAGCTCGACAGCCTTCGCAC 1549
 302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
 1550 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGCGAGAGCCCTTTACATA 1599
 313 313
 1600 GATTACACGCATTTACGGATATCTTACAATGTGTACTTAGCAGCTACTC 1649
 314AspIleAspGlnCys..... 318
 1650 CATTGCCACGCCTTGCAAGATATATATACCTGCTTACCTGGGAGAGGGC 1699
 319SerSerAspGlyTrpPhe 324
 1700 TCCTTCAACATGGCTCCTGTGCAGACATCAAGAAGTTGAGCGCTGGCAG 1749
 325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
 1750 GTCTGTGAAGCACCTACCGCATCTAACTTTACAAACAATATGGGGAGCA 1799
 336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
 1800 GGTGACCTTTGATGAGTGTGGTGACCTGGTGGGGAAGCTATTCCATCATCA 1849
 348Gly 348
 1850 ACTGGCACCTCTCCCCAGAGGATGGCTCCATCGTGTTAGGAAGTCGGG 1899
 349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGlu 362
 1900 TATTACAACTGCTATGCCAAGAAGGGAAGAAAGACTCTTCTATCAACAGGA 1949
 362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
 1950 GAAMATCTGTGGAGTGGGTCTCCAGGAGGAGTCCCTTCTCCAATGCA 1999
 378IleSerGlySerThrLysAspValSerGluGlu..... 388
 2000 GCCGAGACTCGCTGGCAGGACCAGGAAGGATCATTTGAGGGGGAGGCC 2049
 389AlaTyrValCysLeuProCysArgGluGly..... 398
 2050 ACCTGCTGCTTGAGTGTGTGAGTGTCTCGATGATGGGAGTATAGTGATGA 2099
 399CysProPheCysAlaAspAsp..... 405
 2100 GACAGATGCCAGTGCTTAAACAGTCCCGCAGATGACTCTGTGTCCAATG 2149
 406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
 2150 AGAACACACCTCTCGATTGGCCAGGAGATCAGTTCTGTCTGTGGAGC 2199
 418LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
 2200 GAGCCCTTTGGGATCGCACCTCTTTTGGCGTGTCTGGCATTTTCTCT 2249
 430 uAspPheValSerMetLeuValTyrHisPheArgLysAlaLysSerI 447
 2250 GACAGCCTTTGTCTGGGTGTGTTTATCAAGTTTCCGCAACACCATTTG 2299
 447 leArgAlaSerGlyLeuIleLeuGluThrIleLeuPheGlySerLeu 463


```
313 ..... 313
1600 GATTACACGATTACGGATATCCTACAATGTGTACTTAGCAGTCTACTC 1649
314 .....AspIleAspGlnCys..... 318
1650 CATTGCCACCGCCTTGCAGAATATATATACCTGCTTACCTGGGAGAGGGC 1699
319 .....SerSerAspGlyTrpPhe 324
1700 TCTTCACCAATGCTCTCTGCGACACATCAAGAAAGTTGAGGGCTGCGAG 1749
325 SerGlyThrHisLysCysHisLeuAsnSer..... 335
1750 GTCTGAAGCACCTACCGGCATCTAAACTTTACAAACAATATGGGGAGCA 1799
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
1800 GGTGACCTTTGATGAGTGTGGTGACCTGGTGGGGAACATTATCCATCATCA 1849
348 .....Gly 348
1850 ACTGGCACCTCTCCCGACAGAGATGGCTCCATCGTGTTTAAGGAAGTCGGG 1899
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
1900 TATTACAACGCTCTATGCCAAGAAGGGAGAAGACTCTTTCATCAACGAGGA 1949
362 yValLeuProValAsnPheArgArgGlyProAspGlnHis..... 377
1950 GAAATCCTGTGAGTGGTGTCCAGGGAGGTGCCCTTCTCCAACATGCA 1999
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
2000 GCCGAGACTCCCTGGCAGGACCAAGGAAGGATCATTTGAGGGGAGCCCC 2049
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGTGCTTTGAGTGTGTGGAGTGTCTGTATGGGAGGTATAGTGATGA 2099
399 .....CysProPheCysAlaAsp..... 405
2100 GACAGATGCCAGTCCCTGTAAACAGTCCCGCAGATGACTTCTGTGTCCAATG 2149
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
2150 AGAACACACCTCTCGATGTGCCAAGAGATCGAGTTTCTGTCTGTGGACG 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLe 430
2200 GAGCCCTTTGGGATCGCACTCAACCCTTTTCCGCTGTGTGGCATTTTCTCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
2250 GACAGCCTTTGTCTGGTGTGTATCAAGTTCCCGCAACACACCCATG 2299
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
2300 TCAAGGCCACCAACCGAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCTC 2349
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
2350 TGCTGTCTTCTCCAGCTCCCTGTTCTTTCATCGGGAGCGCCGAGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
2400 GTCCCGCTCGCGCCAGCGGCTTTGGCATGACGTTCTGTCTGTCTGTCTGT 2449
497 lyThrValThrLysLeuHisArgValLeuLysValPheLeuSerArg 513
2450 CATGCATCTGTGTGAACCAACCGTGTCTCTGTGTGTGTGTGTGTGTGTGT 2499
```

```
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
2500 ATCCCCACACGAGTTCCACCGCAAGTGTGGGGCTCAACCTGCAGTTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
2550 GCTGGTTTCTCTGACCTTCATGAGATGTGTCATCTGTGTGATCTGGC 2599
540 heLeuIleGlyTrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
2600 TCTACACCGCGCCCTCAAGCTACCGCAACCGAGGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
2650 ATCATCTTCATCAGTCCACGAGGGCTCCTCATGGCCCTGGGCTTCT 2699
568 uIlePheAsnMetCysLeuIle 575
2700 GATCGGCTACACCTGCTGCTG 2721

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:us-08-943-986-3

seq_documentation_block:
; Sequence 3, Application US/08943986
; Patent No. 5962314
; GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,986
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/484,565
; FILING DATE: 7-June-1995
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
```



```

; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 373..3606
; OTHER INFORMATION:
;
; US-08-943-986-3

```

```

alignment_scores:
    Quality: 161.00      Length: 676
    Ratio: 0.583         Gaps: 27
Percent Similarity: 40.828  Percent Identity: 20.118

```

```
alignment_block:

```

```
US-09-775-181-4 x US-08-943-986-3 ..
```

```
Align seg 1/1 to: US-08-943-986-3 from: 1 to: 3809
```

```

75 AlaGluGluValPrometAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   :::::::::::::: :::::
76 TCAGAGCATTCCCTCTACGATTCTGTGGGAGCAACTGGCTCAGG 815
   :::::::::::::: :::::
91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108
   :::::::::::::: :::::
816 CGTCTCCAGCGGAGTGGCAAACTCTGCTGGGCTCTCTACATT..... 858
   :::::::::::::: :::::
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   :::::::::::::: :::::
859 .....CCCCAGGTCAGTTATGCTCTCCAGC.....AGA 888
   :::::::::::::: :::::
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuGl 141
   :::::::::::::: :::::
889 CTCCTCAGCAACAAGAAATCAATTCAAGTCTTCTCCGAAAC...ATCCC 935
   :::::::::::::: :::::
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrG 158
   :::::::::::::: :::::
936 CAATGATGAGCAGCAGGCGCCTGATGCGAGACATCATCGAGATTTCC 985
   :::::::::::::: :::::
158 InAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   :::::::::::::: :::::
986 GCTGGAACCTGGGTGGGCACAATTGTCAGCTGATGACGACTATGGCGCGCG 1035
   :::::::::::::: :::::
173 .....ArgAlaAlaIleThrPheSerT 180
   :::::::::::::: :::::
1036 GGGATTGAGAAATTCGAGAGGAAGCTGAGGAAGAGGATATCTGCATCGA 1085
   :::::::::::::: :::::
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   :::::::::::::: :::::
1086 CTTCACTGAACATCTCTCCAGTACTCTGATGATGAGGAAGAGATCCAGCATG 1135
   :::::::::::::: :::::
196 .....ArgGluGluSerArgIleLeuLeuG 204
   :::::::::::::: :::::
1136 TGGTAGAGTGATTCAAAATTCACGGCCAAAGTCATCGTGGTTTCTCC 1185
   :::::::::::::: :::::
204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   :::::::::::::: :::::
1186 AGTGGCCAGATCTTGAGCCCTCATCAGGAGATTGTCGCGCGCAA... 1232
   :::::::::::::: :::::
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   :::::::::::::: :::::
1233 .....TATCAGGGCAAGATCTGGCTGGCCAGCAGGAGCCTGGGCCAGCT 1276

```

```

237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   :::::::::::::: :::::
1277 CCTCCTGATGCCCATGCTCAGTACTTCCACGTTGGTGGCGGCACCAATT 1326
   :::::::::::::: :::::
250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
   :::::::::::::: :::::
1327 GGATTTCGCTCTCAAGGC.....TGGGCA 1349
   :::::::::::::: :::::
267 ProProTyrLeuGlu.....Cy 272
   :::::::::::::: :::::
1350 GATCCAGGCTTCGGGAATTCCTCAAGAGGTCCATCCAGGAAGTCTG 1399
   :::::::::::::: :::::
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   :::::::::::::: :::::
1400 TCACAATGTTTTCGCAAGGAGTTTGGGAAGAAACATTTAACTGCCAC 1449
   :::::::::::::: :::::
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   :::::::::::::: :::::
1450 CTCCAAGAAGGTGCAAAAGGACCTTTACCTGTGGACACCTTTCTGAGAGG 1499
   :::::::::::::: :::::
295 .....AsnLeuValProGluPheArg 302
   :::::::::::::: :::::
1500 TCACGAAGAAAGTGGGACAGGTTTAGCAACAGCTCGACAGCCTTCCGAC 1549
   :::::::::::::: :::::
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   :::::::::::::: :::::
1550 CCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGAGCCCTTACATA 1599
   :::::::::::::: :::::
313 ..... 313
1600 GATTACAGCATTTACGGATATCTTACAATGTGTACTTAGCAGTCTACTC 1649
   :::::::::::::: :::::
314 .....AspIleAspGlnCys..... 318
   :::::::::::::: :::::
1650 CATTCGCCACGCTTGCAGATATATATACCTGCTTACCTGGGAGAGGCG 1699
   :::::::::::::: :::::
319 .....SerSerAspGlyTrpPhe 324
   :::::::::::::: :::::
1700 TCTTCAACAATGGCTCTGTGCAGACATCAAGAAAGTTGAGCGGTGGCAG 1749
   :::::::::::::: :::::
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   :::::::::::::: :::::
1750 GTCCTGAAGCACCTACGGCATCTAAACTTTACAAACAATATATGGGGAGCA 1799
   :::::::::::::: :::::
336 .....GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
   :::::::::::::: :::::
1800 GGTGACCTTTGATGAGTGTGTGACCTGGTGGGAACATTTCATCATCA 1849
   :::::::::::::: :::::
348 .....Gly 348
   :::::::::::::: :::::
1850 ACTGGCACCTCTCCCGAGAGGATGGCTCCATCGTGTGTTAAGGAAGTCGGG 1899
   :::::::::::::: :::::
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProGl 362
   :::::::::::::: :::::
1900 TATTACAAGCTCTATCCCAAGAGGAGAAAGACTCTTTCATCAACGAGGA 1949
   :::::::::::::: :::::
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   :::::::::::::: :::::
1950 GAAATCTGTGGAGTGGTCTCCAGGAGGTGCCCTTCTCCAACTGCA 1999
   :::::::::::::: :::::
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   :::::::::::::: :::::
2000 GCCGAGACTGCCTGGCAGGACCAAGAGGATCATTTAGGGGGGAGCCC 2049
   :::::::::::::: :::::
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   :::::::::::::: :::::
2050 ACCTGCTGCTTGTAGTGTGGAGTCTCTGATGGGAGTATAGTATGATGA 2099
   :::::::::::::: :::::
399 .....CysProPheCysAlaAspAsp..... 405
   :::::::::::::: :::::
2100 GACAGATGCCAGTGCCTGTAAACAAGTGGCCAGATGACTTCTGGTCCAATG 2149
   :::::::::::::: :::::

```


141 nSerAsnLysSerArgGluInAsnLeuGlnInAspAspLeuAspTrpTyrG 158
::: ::::: ::: ::::: ::::: ||| ::::: |||
936 CAATGATAGCAGCGGCACCTGGCCATGGCAGCATCATCGAGTATTTC 985

158 lnAlaLeuValTrpSerLeuLeuGlyGluProSerIleSer..... 172
::: ::::: ||||::: ||||::: ::::: :::::
986 GCTGGAACTGGGTGGCAATTCAGCTGATGACGACTATGGCGGCCG 1035

173ArgAlaAlaIleThrPheSerT 180
1036 GGGATTGAGAAATTCGAGAGGAAGCTCAGGAAGGATATTCGCATCGA 1085

180 hrAspSerLeuSerAlaProLanValPheLeuGlnAlaThr... 195
||::: ||||::: ||||::: ||||::: ||||
1086 CTTTCAGTGAACTCATCTCCCAGTACTCTGATGAGGAAGATCCAGCATG 1135

196ArgGluGluSer ArgIleLeuLeuG 204
1136 TGGTAGAGTGATTCAAATTCACGGCCAGGATCATCGTGGTATTTCCTC 1185

204 InAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
|| ||| ||||||||| ||||| ||||| |||||
1186 AGTGGCCAGATCTTGAGCCCTCATCAAGAGAGATTGTCCGGCGCAR... 1232

221 GluTrpPheHisGlyLeuArgLysTrpArgProHisLeuHisArgAr 237
::: ||||| ::::: ||||| ||||| |||||
1233TATCAGGGCAAGATCTGCTGGCCAGCGAGCCTGGGCCAGCT 1276

237 gglyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
||| ::::: ||||| ||||| ||||| |||||
1277 CTCTCCCTGATCCCATGCTCAGTACTTCCACGTGGTGGCGGCACCATT 1326

250 rpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266
|| ||| ::::: ||||| ||||| ||||| |||||
1327 GGATTCGCTCTGAAGC.....TGGGCA 1349

267 ProProTyrrLeuGlu.....Cy 272
||| ||| ||||| ||||| |||||
1350 GATCCAGGCTTCGGGAATTCCTCAAGAAGTCCATCCAGGAAGTCGTG 1399

272 s.GluAsnGlySerTyrrLysProGlyTrpLeuValThr..... 284
|| ::||| |||| | ||||| ||||| |||||
1400 TCACAATGGTTTCCCAAGGAGTTTTGGGAAGAAACATTAATACTGCCAC 1449

285 LeuSerSerAlaIleTyrrClyLeuGlnPro..... 294
||| ::::: |||| | |||| | |||| |
1450 CTCCAAGAGGTGCAAAAGGACCTTACCTGTGACACCTTCTGAGAGG 1499

295AsnLeuValProGluPheArgG 302
||| ||||| ||||| |||||
1500 TCACGAAGAAAGTGGCGCACAGTTTAGCAACAGCTCCACAGCCTTCGCAC 1549

302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
::: ||| ||||| ::::: ||||| |||||
1550 CCTCTGTACAGGGATGAGAACAATCAGCAGTGTGAGAGCCCTTACATA 1599

313 313

1600 GATTACAGCATTTACGGATATCTACAATGTGTACTTACGACTTACTC 1649

314AspIleAspGlnCys..... 318
||| ||||| ||||| |||||
1650 CATTCGCCACGGCTTGCAAGATATATACCTGCTTACCTGGGAGAGGCG 1699

319SerSerAspGlyTrpPhe 324
||| ||||| ||||| |||||
1700 TCTTCACCAATGGCTCCTGTGCAGACATCAAGAAAGTTGAGCGGTGCAG 1749

325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
||| ||||| ||||| |||||
1750 GTCTGAAGCAGCTTACGCGCATTAACATTTTACAACAACATATGGGGAGCA 1799

```

336 .....GlucyMetProIleIleLysGlyLeuGlyPheValLeu... 347
      |||||
1800 GGTGACCTTTGATGAGTGTGTGAGCTGGTGGGAACTATTCCATCATCA 1849
      |||||
348 .....Gly 348
      |||
1850 ACTGGCACCTCTCCACAGAGATGGCTTCATCGTTTAAAGNAGTCGGG 1899
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
      |||||
1900 TATTACAACGCTTATGCCAAGAAGGAGAAACACACTCTTTCATCAACGAG 1949
      |||||
362 yValLeuProValAsnAsnPheArgArgArgGlyProAspGlnHis.... 377
      : : : : :
1950 GAAATCCTGTGGAGTGGTGTCTCCAGGAGGTGCCCTTCTCCAACGTGCA 1999
      : : : : :
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
      : : : : :
2000 GCCGAGACTGCTGCAGGGACACAGAAAGGATCATTTGAGGGGGAGCCCC 2049
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
      : : : : :
2050 ACCTGTCTTTGAGTGTGTGGAGTGTCTGATGGGAGTATAGTCATGA 2099
      : : : : :
399 .....CysProPheCysAlaAspAsp..... 405
      |||||
2100 GACAGATGCCAGTGCCTCTAACAAAGTCCACAGTACTTCTGGTCCAATG 2149
      : : : : :
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
      : : : : :
2150 AGAACACACACTCTCTGCATTCGCAGGAGATCGAGTTCTTGTGTGGGAGC 2199
418 .....LeuAlaIleIleSerPheGlnGlyLeuCysMetLeuLeu 430
      : : : : :
2200 GAGCCCTTTGGGATCGCACTCACCTCTTTGCGGTCTGGCATTTTCTCT 2249
      : : : : :
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
      : : : : :
2250 GACAGCCCTTTGTGTGGTGTGTTTATCAAGTTTCGCAACACACACCAT 2299
      : : : : :
447 leArgAlaSerGlyLeuIleLeuLeuGluThrIleLeuPheGlySerLeu 463
      : : : : :
2300 TCAGGGCCACCAACGAGCTCTCCTACCTCCTCTCTCTCTCCCTGCTC 2349
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480
      : : : : :
2350 TGTGCTTCTCCAGCTCCTGTTCTTTCATCGGGGAGCCCGACGATGGAC 2399
480 gCysIleLeuLeuArgTrrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
      |||||
2400 GTGCGCGCTGCGCCAGCGCGCTTTGGCATCAGTCTCGTCTGCTGATCT 2449
      : : : : :
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
      : : : : :
2450 CATGCATCTCGGTGAAACCAACACCGCTGCTCTCTGTTGTGGGCCAAG 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
      : : : : :
2500 ATCCCCACCACTTCCACGCGAAGTGTGGGGCTCAACCTGCAGTTCCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrrp 540
      |||||
2550 GCTGGTTTTCCTCTGCACCTTCATGCAGATTTCTCATCTGTGTGATCTGC 2599
      : : : : :
540 heLeuIleGlyTrrpThrSerSerVal...CysGlnAsnLeuGluLysGln 555
      : : : : :
2600 TCTACACGCGCCGCCCTCAAGCTACCGCAACACGAGGCTGGAGGATGAG 2649
555 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
      : : : : :
2650 ATCATCTTCATCACTGCACGAGGGGCTCCCTCATGTGGCCCTGGGCTTCT 2699
568 uIlePheAsnMetCysLeuIle 575

```

```
|||||
2700 GATCGGCTACACCTGCCTGCTG 2721

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-484-719B-3
seq_documentation_block:
; Sequence 3, Application US/08484719B
; Patent No. 6031003
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth, Edward M.
; APPLICANT: Brown, Steven C. Hebert,
; APPLICANT: Bradford C. Van Wagenen,
; APPLICANT: Manuel F. Balandrin,
; APPLICANT: Forrest H. Fuller, Eric G.
; APPLICANT: Delmar, Scott T. Moe
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS word
; SOFTWARE: FASTSEQ for Windows Version 3.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,719B
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas C. Murdock
; REGISTRATION NUMBER: 37,549
; REFERENCE/DOCKET NUMBER: 213/007
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 373..3606
```

```
US-08-484-719B-3

alignment_scores:
Quality: 161.00 Length: 676
Ratio: 0.583 Gaps: 27
Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block:
US-09-775-181-4 x US-08-484-719B-3 ..
Align seg 1/1 to: US-08-484-719B-3 from: 1 to: 3809

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspse 91
:::|||||:::|||||:::|||||:::|||||:::
766 TCAGAGACACATTCCTCTAGATTGCTGTGGTGGAGCACTGGCTCAGG 815
91 rHisGlnLeuLysArgAlaAsnCyssSerGlyArgTyrGluLeuAlaGlyL 108
: : : : : ||||| : : : : :
816 CGTCTCCACGGCAGTGGCAAAATCTGCTGGGGCTCTTCTACATT..... 858
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
||| : : : : : ||||| : : : : :
859 .....CCCCAGGTCAGTTATGCCCTCTCCACG.....AGA 888
125 AlaLeuAspThrLeuThrHisAlaThrAsnPhelHeAsnValMetLeuG1 141
||| : : : : : : : : : : : : : : :
889 CTCCTCAGCAACAAGAATCAATTCAGTCTTCTCCGAACC...ATCCC 935
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTyrTrpG 158
: : : : : : : : : : : : : : : : :
936 CAATGATGAGCACCAGGCCACTGCCATGGCAGACATCATCGAGTATTTC 985
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
: : : : : ||||| : : : : : : : : : :
986 CTTGGAACTGGTGGGCACAAATTCAGCTGATGACGACTATGGCGCGCG 1035
173 .....ArgAlaAlaIleThrPheSerT 180
||| : : : : : ||||
1036 GGGATTGAGAAATTCGAGAGGAAGCTGAGGAAAGGATATCTGTCATCGA 1085
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
||| : : : : : : : : : : : : : :
1086 CTTCACTGAACATCATCTCCAGTACTCTGATGAGGAAGAGATCCAGCATG 1135
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
||| : : : : : ||||| : : : : :
1136 TGGTAGAGGTGATTCAAATTCACGCCCAAGTCACTCGTGGTTTCTCC 1185
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
||| : : : : : : : : : : : : : :
1186 AGTGGCCAGATCTTGAGCCCTCATCAAGGAGATGTCCGGCGCAA... 1232
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
: : : : : ||||| : : : : :
1233 .....TATCAGCGGCAAGATCTGGCTGGCCAGCGAGGCTGGGCCAGCT 1276
237 gglyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
||| : : : : : ||||| : : : : :
1277 CCTCCCTGATCGCCATGCTCTAGTACTTCCACGTGTTGGCGGCACCAT 1326
250 rpArgArgLysAspGlyLeuGlyArgLysTrpArgProHisLeuHisArg 266
||| : : : : : ||||| : : : : :
1327 GGATTCGCTCTGAAGC.....TGGGCA 1349
267 ProProTyrLeuGlu.....Cy 272
||| : : : : : ||
1350 GATCCCGAGCTTCCGGGAATTCCTGAAGAAGGTCCATCCAGGAGTCTG 1399
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
: : : : : ||||| : : : : :
1400 TCCACAATGGTTTTGCCAAGGAGTTTGGGAAGAAACATTTAACTCCAC 1449
```

285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
|||||:|||||
1450 TCCTCAAGAGGTCGAAAGGACCTTTACCTGTGGACACCTTTCTGAGAGG 1499
295AsnLeuValProGluPheArgG 302
|||:|||||
1500 TCACGAGAAAGTGGCAGACGGTTTAGCAACAGCTCGACAGCCTTCCGAC 1549
302 lyValMetLysValAspLeuGlnLysVal..... 313
:::|||||:|||||
1550 CCTCTGTACAGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1599
313 313
1600 GATTACAGCATTTACGGATATCTTACAAATGTGTACTTAGCAGTCTACTC 1649
314AspLeuGlnCys..... 318
|||||:|||||
1650 CATTCGCCACGCTTGCAGATATATATACCTGTCTACTGGGAGAGGC 1699
319SerSerAspGlyTrpPhe 324
:::|||||:|||||
1700 TCTTCAACAATGGCTCCTGTGTGAGACATCAAGAAAGTTGAGGCGTGGCAG 1749
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
|||:|||||:|||||
1750 GTCTGTGAAGCAGCTACGGCATCTAAACATTTTAAACAATATGGGGAGCA 1799
336GluCysMetProIleLysGlyLeuGlyPheValLeu. 347
|||||:|||||:|||||
1800 CGTGACCTTTGATGAGTGTGGTGTGACCTGTGTGGGAACATTCCCATCATCA 1849
348Gly 348
1850 ACTGGCACCTCTCCCCAGAGGATGCTCCTCATCTGTGTTTAAAGGAAGTCGGG 1899
349 AlaTyrGluCysIleCysIleAlaGly.....PheTyrHisProG1 362
|||||:|||||:|||||
1900 TATTACAACGCTATATGCCAAGAGGAGGAAAGACATCTTTCATCAAGGAGA 1949
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis..... 377
:::|||||:|||||:|||||
1950 GAAATCTGTGGTGGTGTCTCCAGGAGGTGCCCTTCTCCAACTGCA 1999
378IleSerGlySerThrLysAspValSerGlu..... 388
:|||||:|||||:|||||
2000 GCCGAGACTGCTGCGAGGACAGGAAAGGATCATTTGAGGGGAGCCC 2049
389AlaTyrValCysLeuProCysArgGluGly..... 398
2050 ACCTGCTGTTGAGTGTGTGGAGTGTCTCTGATGGGAGTATAGTGATGA 2099
399CysProPheCysAlaAsp..... 405
2100 GACAGATCCAGTGTCTTAACAAGTCCGACAGATCTTGTGGTCCAATG 2149
406SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
:::|||||:|||||:|||||
2150 AGAACACACAGCTCTCTGATGCAAGGAGATCGAGTTCTGTCTGCGGAGC 2199
418LeuAlaIleSerPheGlnGlyLeuCysMetLeuLe 430
:|||||:|||||:|||||
2200 GAGCCTTTGGGATCGCACTACCCCTTTTGGCGTGTGGCATTTTCCT 2249
430 uAspPheValSerMetLeuValValTyrHisPheArgLysAlaLysSerI 447
|:|||||:|||||:|||||
2250 GACACCTTTGTGTGGTGTGTTTATCAAGTTCGCAACACACACCATATG 2299
447 leArgAlaSerGlyLeuIleLeuGluThrIleLeuPheGlySerLeu 463
:|||||:|||||:|||||
2300 TCAAGGCCACCAACGAGAGCTCTCTACCTCTCTCTCTCTCTCTCTCT 2349
464 LeuLeuTyrPheProValIleLeuTyrPheGluProSerThrPheAr 480

2350 TGCTGCTTCTCCAGCTCCCTGTTCTTTCATCGGGGAGCCCCAGGACTGGAC 2399
480 gCysIleLeuLeuArgTrpAlaArgLeuLeuGlyPheAlaThrValTyrG 497
|||:|||||:|||||
2400 GTGCGCGCTGCGCCAGCGGCTTTGGCATCAGCTTCGTGCTCTCATCT 2449
497 lyThrValThrLeuLysLeuHisArgValLeuLysValPheLeuSerArg 513
:::|||||:|||||:|||||
2450 CATGATCTCTGTGTAACCAACACCGCTGTCTCTGTGTGTGGGCAAG 2499
514 ThrAlaGlnArgIleProTyrMetThrGlyGlyArgValMetArgMetLe 530
:::|||||:|||||:|||||
2500 ATCCCCACACAGCTTCCACCGCAAGTGGTGGGGCTCAACCTGCAGTTCT 2549
530 uAlaVal.....IleLeuLeuValValPheTrpP 540
|||:|||||:|||||
2550 GCTGCTTTTCTCTCCACCTTCATCGCATTTGTCATCTGTGTGATCTGGC 2599
540 heLeuIleGlyTrpTrpSerSerVal...CysGlnAsnLeuGluLysGln 555
:::|||||:|||||:|||||
2600 TCTACACGCGCGCCCTCAAGCTACCGCAACGAGGAGCTGGAGGATGAG 2649
556 IleSerLeuIle.....GlyGlnGlyLysThrSerAspHisLe 568
|||:|||||:|||||
2650 ATCATCTTTCATCATCGTCCACGAGGCTCCCTCATGCGCCCTGGGCTTCT 2699
568 uIlePheAsnMetCysLeuIle 575
|||||:|||||:|||||
2700 GATCGGCTACACCTGCCCTGCTG 2721

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-546-998-2

seq_documentation_block:

: Sequence 2, Application US/08546998
: Patent No. 6211244
: GENERAL INFORMATION:
: APPLICANT: Van Wagenen, Bradford C.
: APPLICANT: Moe, Scott T.
: APPLICANT: Balandrin, Manuel F.
: APPLICANT: DelMar, Eric G.
: APPLICANT: Nemeth, Edward F.
: TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: First Interstate World Center
: STREET: Suite 4700
: STREET: 633 West Fifth Street
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: FASTSEQ
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/546,998
: FILING DATE: October 23, 1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: PRIOR APPLICATION DATA: including application
: #APPLICATION NUMBER: described below:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Heber, Sheldon O.
: REGISTRATION NUMBER: 38,179
: REFERENCE/DOCKET NUMBER: 215/304
: TELECOMMUNICATION INFORMATION:

No. 6211244e

```

; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 373...3606
; OTHER INFORMATION:
;
; US-08-546-998-2

```

alignment_scores:

Quality: 161.00 Length: 676
Ratio: 0.583 Gaps: 27
Percent Similarity: 40.828 Percent Identity: 20.118

alignment_block:

US-09-775-181-4 x US-08-546-998-2 ..

Align seg 1/1 to: US-08-546-998-2 from: 1 to: 3809

```

75 AlaGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 TCAGAGCAGATCCCTACGATTGCTGTGGGAGCAACTGCTCAGG 815
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 rHisGlnLeuLysArgAlaAsnCysserGlyArgTyrGluLeuAlaGlyL 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
816 CGTCTCCAGCGCACTGCCAAATCTGCTGGGCTCTTCTACAT..... 858
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
859 .....CCCCAGGTGAGTTATGCCCTCCACG.....AGA 888
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 AlaLeuAspThrLeuThrHisAlaThrAsnPheLeuAsnValMetLeuG1 141
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
889 CTCCTCAGCAACAAGATCAATCAAGTCTTCTCCGAACC....ATCCC 935
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
936 CAATGATGAGCAGCAGCCCACTGCCATGGCAGACATCATCGAGTATTTC 985
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 lnAlaLeuValTrpSerLeuLeuGluGlyGluProSerIleSer..... 172
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
986 GCTGGAACTGGTGGGCACAAATTCGAGCTGATGACGACTATGGGGCGCG 1035
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 .....ArgAlaAlaIleThrPheSerT 180
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1036 GGGATTGAGAAATCCGAGAGGAGCTGAGGAAGGATATCTGCATCGA 1085
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1086 CTTTCAGTGAACATCTCCCACTCTGATGAGGAAGAGATCCAGCATG 1135
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....ArgGluGluSer.ArgIleLeuLeuG 204
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 TGTAGAGGTGATTCAAAATTCACGCGCCAAAGTCATCGTGGTTTCTCC 1185
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuGluThr 220
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1186 AGTGGCCAGATCTTGAGCCCTCATCAGGAGATTGTCCGGCGCAA... 1232
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArgAr 237
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1233 .....TATCAGCGGCAAGATCTGGCTGGCCAGCGAGGCTGGGCCAGCT 1276
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1277 CCTCCCTGATCGGCATCGCTCAGTACTTCCAGCTGGTTGGCGGACCACTT 1326
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 rArgArgLysAspGlyLeuGlyLysSerHisPheLysTrpSer 266
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1327 GGATTCCTCTGAAGGC.....TGGGCA 1349
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ProProTyrLeuGlu.....Cy 272
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1350 GATCCAGGCTTCGGGAATTCCTGAAGAAGTCCATCCAGGAAGTCTG 1399
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1400 TCCACAATGTTTGGCAAGGAGTTTGGGAAGAACAATTAACATGCCAC 1449
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 LeuSerSerAlaIleTyrGlyLeuGlnPro..... 294
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 CTCAGAAGGTGCAAAAGGACCTTTACCTTGGACACACCTTTCTGAGAG 1499
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 .....AsnLeuValProGluPheArg 302
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1500 TCACGAAGAAGTGGCGACAGGTTTAGCAACAGCTCGACAGCCTTCGGAC 1549
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 lyValMetLysValAspIleAsnLeuGlnLysVal..... 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1550 CCCTCTGTACAGGGGATGAGAACATCAGCAGTGTGAGACCCCTTACATA 1599
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
313 ..... 313
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1600 GATTACAGCATTTACGGATATCTCAATCTGTACTTACAGTCTACTC 1649
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 .....AspIleAspGlnCys..... 318
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1650 CATTTGCCACGCCCTTGCACAGATATATATACCTTACCTTGGGAGAGGC 1699
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 .....SerSerAspGlyTrpPhe 324
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1700 TCTTCAACATGGCTCCTGTGCAGACATCAAGAAAGTTGAGCGTGGCAG 1749
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 SerGlyThrHisLysCysHisLeuAsnAsnSer..... 335
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1750 GTCTGAGCACCTACGGCATCTAAACTTTACAAACAATATGGGGAGCA 1799
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 .....GluCysMetProLysGlyLeuGlyPheValLeu. 347
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1800 GGTGACCTTTGATGAGTGTGTGACCTGTGGTGGGAACTATTCCATCATCA 1849
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 .....Gly 348
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1850 ACTGGCACTCTCCCGAGGAGTGGTCCATCGTGTAAAGGAAGTCGGG 1899
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 AlaTyrGluCysIleCysLysAlaGly.....PheTyrHisProG1 362
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1900 TATTACAACGTCTATGCCAAGAGGAGGAAGAACTCTTCAATCAACGAGGA 1949
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHis.... 377
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1950 GAAATCCTGTGAGTGGTGTCTCCAGGGAGGTGCCCTTCTCCAACTGCA 1999
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 .....IleSerGlySerThrLysAspValSerGluGlu..... 388
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2000 GCGGAGACTGCCTGGCAGGAGCAGGAAGAGGATCATTTAGGGGGAGGCC 2049
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 .....AlaTyrValCysLeuProCysArgGluGly..... 398
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2050 ACCTGCTGCTTTGAGTGTGTGAGTGTCTCTGATGGGAGTATAGTATGA 2099
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
399 .....CysProPheCysAlaAsp..... 405
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2100 GACAGATGCCAGTGCCTGTAAACAAGTCCAGATGACTTCTGTGTCCAATG 2149
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 .....SerProCysPheValGlnGluAspLysTyrLeuArg..... 417
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150 AGAACACACCTCTCTCATTTGCCAGGAGATCGAGTTTCTGTGTGGACG 2199
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


1148 ACAATTGAGCTGACGATGACTATGGCGGCCAGGATCGAGAAGTTTCG 1197
83 laSerTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCy 99
1198 AGAGAGCTGAGGAGGGACATCTGATCGATTCAGCAGGCTCAT.. 1245
99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLys 112
1246CTCCCAATACTCTGATGAGGAAAAG 1270
112 rpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
1271 ATCCAGCAGGTGGTGGAGGTGATCCAGAAATCCACGCCAA..... 1311
129 LeuThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLysSe 145
1312 ...AGTCATTTGCT.....CTTCAGCGGCCAGACC 1343
145 ArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrGlnAlaLeuVal 161
1344 TGGAAACCCCTCATAAAGATGCTCGGCCAATATACAGGCGAGATC 1393
162 TrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAla..... 175
1394 TGGCTGCCAGCGAGGCTCGGCCAGCTCTCCCTGATTGCTATGCCCGA 1443
176IleThrPheSerThrAspSerLeuS 184
1444 GTATTTCCATGTGTCGAGGACCATTTGGTTGGTTGAAAGCTGGGC 1493
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluSerArg 200
1494 AGATCCAGGCTTCGGGAATTCCTCGAGAAAGTCCACCCAGCAAGTCT 1543
201 IleLeuLeuGlnAspLeuSerSerAlaProIleLeuAlaAsnAlaTh 217
1544 GTC.....CACATGTTTTGCCAAGGAGTTTGGGAAGAAAC 1581
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
1582 ATTTAACTGCACCTGCAAGAGGTGTAAAGCCCATTTACCGGTGGACA 1631
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
1632 CTTCTCTGAGAGTCCAGAAAGAGGAGTCCAGGTTTAAAGCAACAGTCC 1681
251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266
1682 ACTGCCTCCGACCTCTGTGCACTGGGGAGGAGAACATCAGCAGTGTGCA 1731
266 rProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuV 283
1732 GACTCCTTACATGATATACATATACAGGATATCTACACGCTACT 1781
283 alThrLeuSerSerAlaIleTyrGlyLeuGlnPro.....AsnLeuVal 297
1782 TAGCGGTCTACTCATTTGCTCATGCCCTACAGATATATACACCTGCATA 1831
298 ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValAs 314
1832 CTTGG...AGAGGCTCTTC.....ACCAACGGTTCTCGCGCAGA 1869
314 pIleAspClnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysH 331
1870 TATCAAGAAGTT.....GAAGCTGGCAGGCTCGTAACACACTCGGGC 1913
331 isLeuAsnAsnSer.....GluCys 337
1914 ACCTAAATTTTACCAGCAATATAGGGGAGCAAGTAACCTTCGATGAATGT 1963
338 MetProIleLysGlyLeuGlyPheValLeu..... 347

1964 GGAGACCTGGCAGGAACTATTCCATCATCAACTGGCACCTCTCCCCAGA 2013
348GlyAlaTyrGluCysIleCysL 355
2014 GGACGGCTCCTACTAGTGTTTAAGGAAGTTGGATATTACAATGCTCTATGCCA 2063
355 ysAlaGly.....PheTyrHisProGlyValLeuProValAsnAsn 368
2064 AGAAGCAGAGAGACTCTTTCATCAATGATGAAAAAATCTGTGGAGTGA 2113
369 PheArgArgGlyProAspGlnHis.....IleSerG 380
2114 TTCTCAAGGAGGTGCTTTCTCAACTGCAGTGCAGACTGCCTGGCAG 2163
380 ySerThrLysAspValSerGluGlu.....AlaTyrValCysL 393
2164 GACCAAGAAAGGAATCAATTGAGGGGAGCCCACTGCTGCTTTGAGTGTG 2213
393 euProCysArgGluGly.....Cys 399
2214 TGGAAATGCTCTGATGGGAGTACAGCGACGACAGATGCAAGTGCCTGT 2263
400 ProPheCysAlaAspAsp.....SerProCysPh 409
2264 GATAAGTGCCTCTGATGACTTCTGTCCAATGAGAACCAACACTTCTGCAT 2313
409 eValGlnGluAspLysTyrLeuArg.....LeuAlaI 420
2314 CGCAAGCAGATCGATTTCTGTGGGACCGAGCCCTTCGGGATCGCAC 2363
420 leIleSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeu 436
2364 TCAGGCTCTTTGCTGTCTGGGATTTTCTCACAGCTTCGTCCTGGGC 2413
437 ValValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuII 453
2414 GTTCTCATCAAGTTCGCAACAGCGCCATCGTCAAGGCCACCAACCGGA 2463
453 eLeuLeuGluThrIleLeuPheGlySerLeuLeuTyrPheProValV 470
2464 GCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2513
470 alIleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrp 486
2514 TGTTCCTTCATCGGGAGCCCGACGACTGGACGTCGCCCTTCGCCACCG 2563
487 AlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLe 503
2564 GCCTTTGGCATCAGCTCGTGTCTGTCATCTCTGTCATCTCTGTTGTAAC 2613
503 uHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProT 520
2614 CAATCGGCT 2663
520 yrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuVal 536
2664 GGAAGTGTGGGGCTCAACCTGCAGTTCCTGCTGCTCTCTCTCTCTCT 2713
537 ValPheTrpPheLeuLeuGly.....TrpThrSerSerValCysGlnAs 551
2714 TTCATCAGATTGTCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2763
551 nLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisL 568
2764 GACTACCGCAACACGAGCTG.....GAGACGAGA 2795
568 euIlePheAsnMetCys 573
2796 TCATCTTCATCACCTGC 2812

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-484-565-1

seq_documentation_block:

Sequence 1, Application US/08484565
Patent No. 5763569
GENERAL INFORMATION:
APPLICANT: Edward M. Brown
APPLICANT: Steven C. Hebert
APPLICANT: James E. Garrett, Jr.
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,565
FILING DATE: 7 June, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 9
APPLICATION NUMBER: 08/353,784
FILING DATE: 9 December, 1994
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 213/006
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-484-565-1

alignment_scores: Quality: 152.50 Length: 657
Ratio: 0.524 Gaps: 27

Percent Similarity: 44.292 Percent Identity: 21.005
alignment_block:
US-09-775-181-4 x US-08-484-565-1
Align seg 1/1 to: US-08-484-565-1 from: 1 to: 5275
24 AlaSerArgAspProGln.....GlyArgProAspSe 34
||| :||||| ||| :|||
954 GCTGGGCATCTCCACAGCAGTGGCCAACTGCTGGGGCTCTTACATC 1003
34 rProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGlyA 51
|||||:||||| ||| :|||||
1004 CCCAGGTCAGTATGCT.....CTCCAGCAGACTCCTCAGCAACA 1047
51 rGAlaSerAlaSerAspSerSerAlaProTrp..... 61
|| :||||| ||| :|||||
1048 GAATCAATCAAGTCTCTCCCGCACCATACCCCAATGATGAACACCAGG 1097
62SerArgSerThrAspGlyThrIleLeuAl 71
||| :||||| ||| :|||||
1098 CCACGGCCATGGCTGACATCATCGAGTACTTCGCTGGAAGTGGGTGGGC 1147
71 aGln.....LysLeuAlaGluGluValProMetAspValA 83
|||| :||||| ||| :|||||
1148 ACAATTGACGTGAGATGACTATGGCGCGGATCGAGAGAGTTTCG 1197
83 LaSerTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCy 99
||| ||| ||| :|||||
1198 AGAGGAAGCTGAGGAGGACATCTGCATCGACTTCAGCGAGCTCAT.. 1245
99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLysT 112
1246CTCCCAATCTCTGATGAGGAAAG 1270
112 rProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
||||| :||||| ||| :|||||
1271 ATCCAGCAGGTGGTGGAGGTATCCAGAAATCCACGCCCAA..... 1311
129 LeuThrHisAlaThrAsnPhelAsnValMetLeuGlnSerAsnLysSe 145
:||||| :|||||
1312 ...AGTCATTGCTG.....CTCTCCAGCGGCCAGACC 1343
145 rArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrGlnAlaLeuVal 161
:||||| :||||| ||| :|||||
1344 TGGAAACCCCTCATCAAGAGATCGTCGCGCAATATCACAGGAGGATC 1393
162 TrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAla..... 175
||| :||||| ||| :|||||
1394 TGGTGGCCAGCGAGCTGGCGCAGCTCTTCCTGATTGCTATGCCGA 1443
176IleThrPheSerThrAspSerLeu 184
||| :||||| ||| :|||||
1444 GTATTTCCATGTTGGCGGAGGCACCATTTGGTTGGTTTGAAGCTGGGC 1493
184 eAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
:||||| :||||| ||| :|||||
1494 AGATCCAGGCTTCGGGAATTCCTGCAGAAATCCACCCAGGAAGTCT 1543
201 IleLeuLeuGlnAspLeuSerSerSerAlaProHisLeuAlaAsnAlaTh 217
:||||| :||||| ||| :|||||
1544 GTC.....CACAAATGGTTTTCGCAAGGAGTTTGGGAAGAAAC 1581
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
:||||| :||||| ||| :|||||
1582 ATTTAACTGCCACCTGCAAGAGGTGCTAAAGGCCCATACCGGTGGACA 1631
234 euHisArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
||||| :||||| ||| :|||||
1632 CCTTCTCAGAGGTCCACGAAGAAGGAGTGCAGGTTAAGCAACACGCC 1681
251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266
||| :||||| ||| :|||||

```

503 uHisArgValLeuLysValPheLeuSerArgThrAlaGlnArlleProt 520  

    ::::||||| | | ||||| :::::  

2614 CAATCGGTCTCCTGGTGTGGCCCAAGATTCCCACGCTTCACC 2663  

    :  

520 yMetThrGlyGlyArgValMetMetLeuAlaValIleLeuLeuVal 536  

    ||| ::::||||| | | ||||| :::  

2664 GGAAGTGTTGGGGCTCAACTGCAGTTCTGCTGGTCTCTCCCTGCACC 2713  

    :  

537 ValPheTrpPheLeuIleGly.....TrpThrSerSerValcysGlnAs 551  

    ::: ::::||||| | | ::::||||| ::  

2714 TTCAATGCAGATTTCATCTGTCGCCATTTGGCTCAATAACAGCCCCCCTC 2763  

    :  

551 nLeuGlutylserLeuIleGlyGlnGlyLysThrSerAspHilS 568  

    :::: ::::||||| | | ::::||||| ::  

2764 GAGTACCGCAACCAACGAGCTG.....GAGGACGAGA 2795  

    :  

568 eullePheasnMetCys 573  

    ::::||||| | |  

2796 TCATCTTCATCACCTGC 2812  

    :  

seq_name: /cgcn2_6/ptodata/2/inna/5B_COMB.seq:US-08-480-751-1  

seq_documentation_block:  

; Sequence 1, Application US/08480751  

; Patent No. 5858684  

; GENERAL INFORMATION:  

; APPLICANT: Edward F. Nemeth  

; APPLICANT: Edward M. Brown  

; APPLICANT: Steven C. Hebert  

; APPLICANT: Forrest H. Fuller  

; APPLICANT: James E. Garrett, Jr.  

; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  

; MOLECULES  

; NUMBER OF SEQUENCES: 20  

; CORRESPONDENCE ADDRESS:  

; ADDRESSEE: Lyon & Lyon  

; STREET: First Interstate World Center  

; SUITE: Suite 4700  

; CITY: Los Angeles  

; STATE: California  

; COUNTRY: USA  

; ZIP: 90071  

; COMPUTER READABLE FORM:  

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  

; COMPUTER: IBM PC compatible  

; OPERATING SYSTEM: PC-DOS/MS-DOS  

; SOFTWARE: FASTSEQ  

; CURRENT APPLICATION DATA:  

; APPLICATION NUMBER: US/08/480,751  

; FILING DATE: 7 June, 1995  

; CLASSIFICATION: 435  

; PRIOR APPLICATION DATA:  

; PRIOR APPLICATION DATA: described below: 9  

; APPLICATION NUMBER: 08/353,784  

; FILING DATE: 9 December, 1994  

; APPLICATION NUMBER: PCT/US/94/12117  

; FILING DATE: 21 October, 1994  

; APPLICATION NUMBER: U.S. 08/292,827  

; FILING DATE: 23 August, 1994  

; APPLICATION NUMBER: U.S. 08/141,248  

; FILING DATE: 22 October, 1993  

; APPLICATION NUMBER: U.S. 08/009,389  

; FILING DATE: 23 February, 1993  

; APPLICATION NUMBER: U.S. 08/017,127  

; FILING DATE: 12 February, 1993  

; APPLICATION NUMBER: U.S. 07/934,161  

; FILING DATE: 21 August, 1992  

; APPLICATION NUMBER: U.S. 07/834,044  

; FILING DATE: 11 February, 1992  

; APPLICATION NUMBER: U.S. 07/749,451
```

```

; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 515..3769
; OTHER INFORMATION:
US-08-480-751-1

```

```

Alignment_scores:
  Quality: 152.50      Length: 657
  Ratio: 0.524        Gaps: 27
Percent Similarity: 44.292 Percent Identity: 21.005

alignment_block:
US-09-775-181-4 x US-08-480-751-1 ..

Align seg 1/1 to: US-08-480-751-1 from: 1 to: 5275

24 AlaSerArgAspProGln.....SerArgSerThrAspGlyThrIleLeuAl 71
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
954 GCTCGGCATCTCCACAGAGTGCACACAGTGCACACCTGCTGGGCTCTTCTACATC 1003

34 rProGluArgThrProLysGlyLysProHisAlaGlnProGlyA 51
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1004 CCCAGGTGAGTATGCTCT.....CTCCAGCAGACTCCTCAGCAACAA 1047

51 rGAlaSerAlaSerAspSerAlaProTrp..... 61
   || : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1048 GAATCAATTCAGTCTCTCTCCGACCATATCCCAATGATGACACACAGG 1097

62 ..... 61
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1098 CCACGGCATGCTGACATCATCGAGTACTTCCGCTGGAAGTGGTGGC 1147

71 aGln.....LysLeuAlaGluValProMetAspValA 83
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1148 ACAATTGCACTGACGATGACTATGCGCGGCAGGATCGAGAAGTTTCG 1197

83 laSerTyrLeuTyrThrGlyAspSerHisGlnLeuLysArgAlaAsnCy 99
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1198 AGAGGAAGCTGAGGAGAGGACATCTGCTGACTTCAGCGAGCTCAT... 1245

99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLysT 112
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1246 .....CTCCATACTCTGATCAGGAAAG 1270

112 rpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1271 ATCCAGCAGGTGGTGGAGTGATCCAGAATTCACCGCCAA..... 1311

129 LeuThrHisAlaThrAsnPhelLeuAsnValMetLeuGlnSerAsnLysSe 145
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1312 ...AGTCATGTGCT.....CTTCCAGCGGCCAGAC 1343

145 rArgGluGlnAsnLeuGlnAspLeuAspTrpTyrGlnAlaLeuVal 161
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1344 TGGACCCCTCATCAAGAGATCGTCCGCGGCAATATCACAGGAGGATC 1393

```

```

162 TrpSerLeuLeuGluGlyGluProSerIleSerArgAlaAla..... 175
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1394 TGGCTGGCCAGCGAGGCTGGCCAGCTCTTCCCTGATTGCTATGCCGA 1443

176 .....IleThrPheSerThrAspSerLeuS 184
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1444 GTATTTCCATGTGGTCGAGGACCATTTGGTTTGGTTTGAANGCTGGC 1493

184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgLysSerArg 200
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1494 AGATCCCAGGCTCGGGAATTCCTGCAGAAAGTCCACCCAGGAAGTCT 1543

201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1544 GTC.....CACAAATGGTTTTCGCAAGGAGTTTGGGAAGAAAC 1581

217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1582 ATTTAACTGCCACCTGCAAGAGGTGCTAAAGGCCCATTTACCGGTGGACA 1631

234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1632 CTTCTGAGAGGTCCAGAAAGAGGAGTGCAGGTTAAGCAACAGTCC 1681

251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1682 ACTGCTTCCGACCTCTGTGCATGGGAGGAGAACATCAGCAGTGTGCA 1731

266 rProProTyrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuV 283
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1732 GACTCTTACATGGATTATACACATTTACGGATATCCTACAACGCTCTACT 1781

283 alThrLeuSerSerAlaIleTyrGlyLeuGlnPro.....AsnLeuVal 297
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1782 TAGCGCTTACTCCATTCCTCATGCTTACAGATATATACACCTGCATA 1831

298 ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1832 CTTGG...AGAGGCTCTTC.....ACCAACGGTTCTCTGGCGAGA 1869

314 pIleAspGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysH 331
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1870 TATCAAGAGGTT.....GAAGCTTGGCAGGTCTGAAACACCTGCGGC 1913

331 isLeuAsnAsnSer.....GluCys 337
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1914 ACCTAAATTTTACCAGCAATATGGGGAGCAAGTAACCTTTCGATCAATGT 1963

338 MetProIleLysGlyLeuGlyPheValLeu..... 347
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
1964 GGAGACCTGGCAGGGAACATATTTCCATCATCACTGGCACCTCTCCCCAGA 2013

348 .....GlyAlaTyrGluCysIleCysL 355
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
2014 GGACGGCTCCATAGTGTATTAGGAAGTTGGATATTACAATGTCTATGCCA 2063

355 ysAlaGly.....PheTyrHisProGlyValLeuProValAsnAsn 368
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
2064 AGAAGGAGAGAGACTCTTCATCAATGATGAAAAATTTCTGTGGAGTGA 2113

369 PheArgArgGlyProAspGlnHis.....IleSerGl 380
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
2114 TTCTCAAGGAGGTGCTTCTTCCAACTGCAGTCGAGACTGCCTGGCAG 2163

380 ySerThrLysaspValSerGluGlu.....AlaTyrValCysL 393
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
2164 GACCAGAAAGGAATCATTTAGGGGAGGCCACCTGCTGCTTTGAGTGTG 2213

393 euProCysArgGluGly.....Cys 399
   ||| : : : : ||| : : : : ||| : : : : ||| : : : : ||| : : : : |||
2214 TGGATGTCTTGTATGGGAGTACACCGAGGACAGATGCAAGTCCCTGT 2263

400 PropheCysAlaAspAsp.....SerProCysPh 409

```

```
2264 GATGAAGTGCCTGATGACTTCTGCTCAATGAGAACACACATCTCTCCAT 2313
2409 eValGlnGluAspLysTyrLeuArg.....LeuAlaI 420
2314 CGCAAGAGATCGAGTTCTGCTGGACCGAGCCCTTCGGGATCGCAC 2363
420 leIleSerPheGlnGlyLeuCyMetLeuLeuAspPheValSerMetLeu 436
2364 TCAGGCTCTTTGCTGTGCTGGGATTTCTCACAGCCTTCGTGCTGGC 2413
437 ValValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuI 453
2414 GTCTTTCATCAAGTTCCGCAACACGCCCATCGTCAAGGCCACCAACCGGA 2463
453 eLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProVal 470
2464 GCCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 2513
470 alIleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrp 486
2514 TGTCTTTCATCGGGAGCCCGACGACTGCGCGCTTCGCGCCAGCGC 2563
487 AlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLe 503
2564 GCCTTTGGCATCAGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2613
503 uHisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIlePro 520
2614 CAATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 2663
520 yrMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuLeuVal 536
2664 GGAAGTGGTGGGGCTCAACCTGCAGTTCCTGCTGCTGCTGCTGCTGCT 2713
537 ValPheTrpPheLeuIleGly.....TrpThrSerValCysGlnAs 551
2714 TTCATCGAGATTCTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2763
551 nLeuGluLysGlnIleSerLeuIleGlyGlnGlyLysThrSerAspHisL 568
2764 GAGCTACCGCAACCCAGGCTG.....GAGGACGAGA 2795
568 eullePheAsnMetCys 573
2796 TCATCTTCATCACTGC 2812
```

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-943-986-1

seq_documentation_block:
; Sequence 1, Application US/08943986
; Patent No. 5962314

GENERAL INFORMATION:
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,986
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/484,565
; FILING DATE: 7-June-1995
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 515..3769
; OTHER INFORMATION:
; US-08-943-986-1

alignment_scores:
Quality: 152.50 Length: 657
Ratio: 0.524 Gaps: 27
Percent Similarity: 44.292 Percent Identity: 21.005

alignment_block:

US-09-775-181-4 x US-08-943-986-1 ..
Align seg 1/1 to: US-08-943-986-1 from: 1 to: 5275
24 AlaserArgAspProGln.....GlyArgProaspse 34
954 GCTCGGGCATCTCCACGAGTGGCCCAACCTGCTGGGCTCTTCTACATC 1003
34 rProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGlyA 51
1004 CCCAGGTGAGTATGCT.....CCTCAGCAGACTCTCTCAGCAACA 1047
51 rGAlaSerAlaSerAspSerSerAlaProTrp..... 61
1048 GAATCAATTCAAGTCTCTCTCGCACCACCATCCCAATGATGAACACCAG 1097
62SerArgSerThrAspGlyThrIleLeuAl 71
1098 CCACGGCCCATGGCTGACATCATCGAGTACTTCCGCTGGAACTGGTGGGC 1147

seq_documentation_block:
Sequence 1, Application US/08353784
Patent No. 6011068
GENERAL INFORMATION:
APPLICANT: Edward F. Nemeth, Edward M.
APPLICANT: Brown, Steven C. Hebert, Manuel
APPLICANT: Bradford C. Van Wagenen, Manuel
APPLICANT: F. Balandrin, Forrest H. Fuller,
APPLICANT: Eric G. DelMar, and Scott T. Moe
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
MOLECULES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: First Interstate World Center
STREET: Suite 4700
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,784
FILING DATE: 9 December, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 8
APPLICATION NUMBER: PCT/US/94/12117
FILING DATE: 21 October, 1994
APPLICATION NUMBER: U.S. 08/292,827
FILING DATE: 23 August, 1994
APPLICATION NUMBER: U.S. 08/141,248
FILING DATE: 22 October, 1993
APPLICATION NUMBER: U.S. 08/009,389
FILING DATE: 23 February, 1993
APPLICATION NUMBER: U.S. 08/017,127
FILING DATE: 12 February, 1993
APPLICATION NUMBER: U.S. 07/934,161
FILING DATE: 21 August, 1992
APPLICATION NUMBER: U.S. 07/834,044
FILING DATE: 11 February, 1992
APPLICATION NUMBER: U.S. 07/749,451
FILING DATE: 23 August, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 209/069
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5275 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 515..3769
OTHER INFORMATION:
US-08-353-784-1

alignment_scores: Quality: 152.50 Length: 657

Ratio: 0.524 Gaps: 27
Percent Similarity: 44.292 Percent Identity: 21.005
alignment_block:
US-09-775-181-4 x US-08-353-784-1 ..
Align seg 1/1 to: US-08-353-784-1 from: 1 to: 5275
24 AlaSerArgAspProGln.....GlyArgProAspSe 34
||| :|||
954 GCTCGGGCATCTCCACAGCAGTGGCCCAACTGCTGGGGCTTCTTACATC 1003
34 rProArgGluArgThrProLysGlyLysProHisAlaGlnGlnProGlyA 51
||||| :|||
1004 CCCAGGTACGCTATGCCT.....CTCCAGCAGACTCTCTCAGCAACA 1047
51 rGAlaSerAlaSerAspSerAlaProTrp..... 61
||| :|||
1048 GAATCAATCAAGTCCTTCTCCGACCATATCCCAATGATGACACACAGG 1097
62SerArgSerThrAspGlyThrIleLeuAl 71
||| :|||
1098 CCACGGCCATGGCTGACATCATCGACTACTTCGCTGGAACCTGGTGGC 1147
71 aGln.....LysLeuAlaGluValProMetAspValA 83
||| :|||
1148 ACAATTGCAGCTGACGATGACTATGCGCGCCAGGATCGAGAAGTTTCG 1197
83 laSerTyrLeuTyrThrGlyAspSer.HisGlnLeuLysArgAlaAsnCy 99
||| :|||
1198 AGAGGAAGCTGAGGAGGAGCATCTGCATCGACTTCAGCGAGCTCAT.. 1245
99 sSerGlyArgTyrGluLeuAlaGlyLeuPro.....GlyLysT 112
||| :|||
1246CTCCCAATACTCTGATGAGGAAAG 1270
112 rpProAlaLeuAlaSerAlaHisProSerLeuHisArgAlaLeuAspThr 128
||||| :|||
1271 ATCCAGCAGGTGCTGGAGTGATCCAGAAATTCACCGCCAA..... 1311
129 LeuThrHisAlaThrAsnPheLeuAsnValMetLeuGlnSerAsnLysSe 145
:||||| :|||
1312 ...AGTCATGTCTGT.....CTTCTCCAGCGCCGACAGC 1343
145 rArgGluGlnAsnLeuGlnAspAspLeuAspTrpTyrGlnAlaLeuVal 161
:||| :|||
1344 TGGACCCCTCATCAAGAGATCGTCCGGCGCAATATCACAGCAGGATC 1393
162 TrpSerLeuGluGlyGluProSerIleSerArgAlaAla..... 175
||| :|||
1394 TGGCTGGCCAGCAGGCGCTGGCCAGCTCTTCCCTGATTGCTATGCCGA 1443
176IleThrPheSerThrAspSerLeuS 184
||| :|||
1444 GTATTTCCATGTGTCGAGGCGACCATTTGGTTTGTGAAAGCTGGGC 1493
184 rAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGluSerArg 200
:||| :|||
1494 AGATCCCGAGCTTCCGGGAATCTCTCAGAAAGTCCACCCCGAGGAGTCT 1543
201 IleLeuGlnAspLeuSerSerAlaProHisLeuAlaAsnAlaTh 217
||| :|||
1544 GTC.....CACAAATGTTTTGCCAAGGAGTGTGGGAAGAAAC 1581
217 rLeuGluThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
:||||| :|||
1582 ATTTAACTGCCACCTGCAAGAGGCTGCTAAAGGCCCATTTACCGGTGGACA 1631
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
||||| :|||
1632 CCTCTCTGAGAGGTACGAGAGGAGGAGGTTCCAGAGTTAAGCAACAGTCCC 1681
251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPheLysTrpSe 266

```

1682 ACTGCTTCCGACCTCTGTGGCACTGGGAGGAGACATACAGAGTGTGGA 1731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 rProPofrYrLeuGluCysGluAsnGlySerTyrLysProGlyTrpLeuV 283
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1732 GACTCTTTACATGGATTATACACATTTACGGATATCTCTACACAGCTACT 1781
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 alThrLeuSerSerAlaIleTyrGlyLeuGlnPro.....AsnLeuVal 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1782 TAGCGCTCTACTCCATCTCTCATGCCCTACCAAGATATATACACCTGCATA 1831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValAs 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1832 CTGGG...AGAGGCTCTTC.....ACCACGGTTCTCTCGGCAGA 1869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 pileAspGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysH 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1870 TATCAAGAAGGTT.....GAAGCTTGCAGGTCTCTGAACACCTGCCGC 1913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 IsLeuAsnAsnSer.....GluCys 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1914 ACCTAAATTTTACCAGCAATATGGGGAGCAAGTAACCTTTTCGATGAATGT 1963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 MetProIleLysGlyLeuGlyPheValLeu..... 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1964 GGAGACCTGGCAGGCAATATTCCATCATCACTACAGTGGCACCTCTCCCCAGA 2013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
348 .....GlyAlaTyrGluCysIleCysL 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2014 GCACGGCTCCATAGTGTTTAAGGAAGTTGGATATTACAATGTCTATGCCA 2063
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 ysAlaGly.....PheTyrHisProGlyValLeuProValAsnAsn 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2064 AGAAAGGAGAGAGACTCTTCATCATGATGAAAAAATTCTGTGGAGTGA 2113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 pheArgATgATgGlyProAspGlnHis.....IleSerGl 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2114 TTCTCAAGGGAGGTGCTTTCTCAACTGCAGTCGAGACTGCTGGCAGG 2163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
380 yserThrLysAspValSerGluGlu.....AlaTyrValCysL 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2164 CACCAGGAAGGAATCATTTAGGGGGAGCCCACTGCTGCTTTGAGTGTG 2213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 euProCysArgGluGly.....Cys 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2214 TGGATGTCTGATGGGAGTACAGCGAGACAGATGCAAGTGCCTGT 2263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 ProPheCysAlaAspAsp.....SerProCysPh 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2264 CATAAGTCCCTGATGATGCTTCTGTTCATCAGAACCAACACTCTCTGCAT 2313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
409 eValGlnGluAspLysTyrLeuArg.....LeuAlaI 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2314 CCCCAGAGAGATCGATTTCTGTGTGACCGACGCCCTTCGGGATCGGCAC 2363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 IleLeSerPheGlnGlyLeuCysMetLeuLeuAspPheValSerMetLeu 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2364 TCACGCTCTTTGCTGTGCTGGCATTTTCTCCACAGCCCTTCGTGCTGGGC 2413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 ValValTyrHisPheArgLysAlaLysSerIleArgAlaSerGlyLeuI 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2414 GTCTTTCAATCAAGTTCCGCAACAGCCCATCTCAAGGCGCACCAACCGGGA 2463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 eLeuLeuGluThrIleLeuPheGlySerLeuLeuLeuTyrPheProValV 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2464 CCTCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
470 alIleLeuTyrPheGluProSerThrPheArgCysIleLeuLeuArgTrp 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2514 TGTCTTCTATCGGGAGCCCGCAGGACTGGACGTGCCGCTCGCCAGCCG 2563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
487 AlaArgLeuLeuGlyPheAlaThrValTyrGlyThrValThrLeuLysLe 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

; FILING DATE: 23 August, 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Douglas C. Murdock
 ; REGISTRATION NUMBER: 37, 549
 ; REFERENCE/DOCKET NUMBER: 213/007
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ;

```

alignment_scores:
  Quality: 152.50      Length: 657
  Ratio: 0.524        Gaps: 27
  Percent Similarity: 44.292  Percent Identity: 21.005

alignment_block:
  US-09-775-181-4 x US-08-484-719B-1  ..

Align seg 1/1 to: US-08-484-719B-1 from: 1 to: 5275

24 AlaSerArgAspProGln.....GlyArgProAspSe 34
   GGTCTGGGGTCTCCACAGCAGGTGGCAACCTGCTGGGGCTTCTTCATC 1003

```

1394 TGGCTGGCCAGCAGCGCCTGGGCCAGCTCTTCCCTGATGTACTATGCCGA 1443
176 |||::: ||| ||| |||
1444 GTATTTCNATGTGTGGAGGCACCATTTGGTTTGTTGTAAGCTGGGC 1493
184 erAlaProAlaProGlnValPheLeuGlnAlaThrArgGluGlusSerArg 200
1494 AGATCCCAGGCTTCGGCAATTCCTGCAGAAAGTCCACCACCGAAGTCT 1543
201 IleLeuLeuGlnAspLeuSerSerAlaProHisLeuAlaAlaTh 217
1544 GTC CACAATGGTTTTGCCAAGAGAGTTTGGGAAGAACA 1581
217 rIeuGlnThrGluTrpPheHisGlyLeuArgArgLysTrpArgProHisL 234
1582 ATTTAACCTGGCCACTCCAAGAGGTGTAAAGGCCCATTTACCGGTGGACA 1631
234 euHisArgArgGlyProAsnGlnGlyProArgGlyLeuGlyHisSerTrp 250
1632 CCTTCTGAGAGGTACCAAGAAGAGGTGCCAGTTTAAACAACAGTCCC 1681
251 ArgArgLysAspGlyLeu...GlyGlyAspLysSerHisPhelysTrpSe 266
1682 ACTGCTCTCCGACCTCTGTGCACCTGGGAGGAGAACATCACAGTGTGCA 1731
266 rProProTrpLeuGluCysGluAsnGlySerTrpLysProGlyTrpLeuV 283
1732 GACTCTTTACATGGATTATACACATTTACGGATATCTCAACAGCTCTACT 1781
283 alThrLeuSerSerAlaIleTrpGlyLeuGlnPro.....AsnLeuVal 297
1782 TAGCGCTCTACTCCATTCCTCATCGCTCACAGATATATACACCTGCATA 1831
298 ProGluPheArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
1832 CTCTGG...AGAGGCTCTTC.....ACCAACGGTTCTCTGCCGACA 1869
314 pIleAspGlnCysSerSeraspGlyTrpPheSerGlyThrHisLysCysH 331
1870 TATCAAGAAGTT.....GAAGCTTGGCAGGTCTGTAACACCTCGCGC 1913
331 IsLeuAsnAsnSer.....GluCys 337
1914 ACCTAAATTTACCAGCAATATAGGGGAGCAGTAACATTTCGATGATGT 1963
338 MetProIleLysGlyLeuGlyPheValLeu..... 347
1964 GGAGACCTGGCAGGGAACATTTCATCATCACTGGCACCTCTCCCCAGA 2013
348GlyAlaTrpGluCysIleCysL 355
2014 GGACGGCTCCATAGTGTTAAGGAAGTGGATATATCAATGTCTATGCCA 2063
355 ysAlaGly.....PheTrpHisProGlyValLeuProValAsnAsn 368
2064 AGAAGGAGACAGACTCTTCATCAATGATCAAAAATTTCTGTGGAGTGA 2113
369 PheArgArgArgGlyProAspGlnHis.....IleSerGI 380
2114 TTCTCAAGGGAGTGCCTTTCTCCAACTGCAGTCGAGACTGCCTGGCAGS 2163
380 ySerThrLysAspValSerGluGlu.....AlaTrpValCysL 393
2164 GACCAGGAAGAATCATTTGAGGGGGAGCCCCACTGCTGCTTTGATGTG 2213
393 euProCysArgGluGly.....Cys 399
2214 TGGATGTCTGATGGGAGTACGCGACGACAGATGCAAGTGCCTGT 2263
400 PropHeCysAlaAspAsp.....SerProCysPh 409

2751 CATCAGTGCACGAGGCTCCTCATGCGCTGGCTTCTGATCGGCT 2800

571 smMetCysLeuIle 575

|||||:::

2801 ACACCTGCGTCTG 2814

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq.us-08-484-565-2

seq_documentation_block:

; Sequence 2, Application US/08484565

; Patent No. 5763569

; GENERAL INFORMATION:

; APPLICANT: Edward M. Brown

; APPLICANT: Steven C. Hebert

; APPLICANT: James E. Garrett, Jr.

; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

; TITLE OF INVENTION: MOLECULES

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: First Interstate World Center

; STREET: Suite 4700

; STREET: 633 West Fifth Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: FASTSEQ

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,565

; FILING DATE: 7 June, 1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA: including application

; PRIOR APPLICATION DATA: described below: 9

; APPLICATION NUMBER: 08/353,784

; FILING DATE: 9 December, 1994

; APPLICATION NUMBER: PCT/US/94/12117

; FILING DATE: 21 October, 1994

; APPLICATION NUMBER: U.S. 08/292,827

; FILING DATE: 23 August, 1994

; APPLICATION NUMBER: U.S. 08/141,248

; FILING DATE: 22 October, 1993

; APPLICATION NUMBER: U.S. 08/009,389

; FILING DATE: 23 February, 1993

; APPLICATION NUMBER: U.S. 08/017,127

; FILING DATE: 12 February, 1993

; APPLICATION NUMBER: U.S. 07/934,161

; FILING DATE: 21 August, 1992

; APPLICATION NUMBER: U.S. 07/834,044

; FILING DATE: 11 February, 1992

; APPLICATION NUMBER: U.S. 07/749,451

; FILING DATE: 23 August, 1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Heber, Sheldon O.

; REGISTRATION NUMBER: 38,179

; REFERENCE/DOCKET NUMBER: 213/006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5006 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; FEATURE:

; NAME/KEY: CDS
; LOCATION: 436...3699
; OTHER INFORMATION:
; US-08-484-565-2

alignment_scores:

Quality: 150.00 Length: 690

Ratio: 0.543 Gaps: 27

Percent Similarity: 40.000 Percent Identity: 19.565

alignment_block:

US-09-775-181-4 x US-08-484-565-2 ..

Align seg 1/1 to: US-08-484-565-2 from: 1 to: 5006

75 AlaGluGluValProMetAspValAlaSerTyrLeuTyrThrGlyAspSe 91

829 TCAGAGCACATTCCTCTACGATTGCTGTGGGAGCAACTGGCTCAGG 878

91 rHisGlnLeuLysArgAlaAsnCysSerGlyArgTyrGluLeuAlaGlyL 108

879 CGTCTCCACGCGCAGTGGCAAAATCTGCTGGGCTCTTCTACATT..... 921

108 euProGlyLysTrpProAlaLeuAlaSerAlaHisProSerLeuHisArg 124

922CCCCAGGTCAGTTATGCTCTCTCCAGC.....AGA 951

125 AlaLeuAspThrLeuThrHisAlaThrAsnPhelLeuAsnValMetLeuG1 141

952 CTCCTCAGCAACAAGATCAATTCAGTCTTCTCCGGAACC...ATCCC 998

141 nSerAsnLysSerArgGluGlnAsnLeuGlnAspLeuAspTrpTyrG 158

999 CAATGATGAGCACCAGCCACTGCCATGGCAGACATCATCGAGTATTTC 1048

158 lnAlaLeuValTrpSerLeuLeuGluGluProSerIleSer..... 172

1049 GCTGGAACTGGTGGGCAATTCGAGCTGATGACGACTATGGGCGCGC 1098

173ArgAlaAlaIleThrPheSerT 180

1099 GGGATTGAGAAATCCGAGAGGAGCTGAGGAAGGATATCTGCATCGA 1148

180 hrAspSerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThr... 195

1149 CTTCACTCACTCATCTCCAGTACTCTGATCAGGAAGAGATCCAGCATG 1198

196ArgGluGluSer.ArgIleLeuLeuG 204

1199 TGGTAGAGGTGATTCAAAATTCACGCCCAAGTCATCGTGTTCCTCC 1248

204 lnAspLeuSerSerAlaProHisLeuAlaAsnAlaThrLeuLuthr 220

1249 AGTGGCCCGAGATCTTGAGCCCTCATCAAGGAGATTGCCGCGCA... 1295

221 GluTrpPheHisGlyLeuArgArgLysTrpArgProHisLeuHisArg 237

1296TATCAGCGGCAAGATCTGGCTGGCCAGCGAGCGCTGGCCAGCT 1339

237 gGlyProAsnGlnGly.....ProArgGlyLeuGlyHisSerT 250

1340 CTTCCCTGATGCCATGCTCAGTACTTCCACGTGGTGGGCGCACCAT 1389

250 rpArgArgLysaspGlyLeuGlyGlyAspLysSerHisPheLysTrpSer 266

1340 GGATTCGCTCTGAGGC.....TGGGCA 1412

267 ProProTyrLeuGlu.....Cy 272

1413 GATCCAGGCTTCCGGGAATTCCTCAAGAAGGTCCATCCAGGAAGTCTG 1462

272 s.GluAsnGlySerTyrLysProGlyTrpLeuValThr..... 284

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:15:11 ; Search time 40.53 Seconds
(without alignments)
1125.797 Million cell updates/sec

Title: US-09-775-181-4
Perfect score: 3199
Sequence: 1 MCAMAYPLLLCLLLAQLGLG.....YMTAVGMWSLVSYDGLTIFQ 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 68:*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----------|--------------------|
| 1 | 136 | 4.3 | 999 | 2 T27628 | hypothetical prote |
| 2 | 133.5 | 4.2 | 1079 | 2 I59362 | calcium/polyvalent |
| 3 | 127 | 4.0 | 1078 | 2 A56715 | calcium receptor (|
| 4 | 123 | 3.8 | 1085 | 2 S40476 | Ca(2+)-sensing rec |
| 5 | 120.5 | 3.8 | 879 | 2 JC7160 | metabotropic gluta |
| 6 | 119.5 | 3.7 | 1088 | 2 B56715 | calcium receptor (|
| 7 | 116 | 3.6 | 872 | 2 JH0361 | metabotropic gluta |
| 8 | 115 | 3.6 | 1199 | 2 A41939 | G protein-coupled |
| 9 | 114 | 3.6 | 1180 | 2 JC2132 | metabotropic gluta |
| 10 | 114 | 3.6 | 1212 | 2 JC2131 | metabotropic gluta |
| 11 | 112 | 3.5 | 879 | 2 JH0562 | metabotropic gluta |
| 12 | 111 | 3.5 | 1171 | 2 A42916 | metabotropic gluta |
| 13 | 110.5 | 3.5 | 3507 | 2 T34513 | hypothetical prote |
| 14 | 109.5 | 3.4 | 868 | 2 JC5701 | Erbb kinase activa |
| 15 | 107.5 | 3.4 | 2437 | 2 S53611 | MTBP1 protein - ra |
| 16 | 106.5 | 3.3 | 850 | 2 JC5700 | Erbb kinase activa |
| 17 | 106.5 | 3.3 | 860 | 2 JC5702 | Erbb kinase activa |
| 18 | 106 | 3.3 | 383 | 2 S53716 | delta-like homotei |
| 19 | 106 | 3.3 | 551 | 2 T30806 | metabotropic gluta |
| 20 | 104.5 | 3.3 | 644 | 1 A40212 | uromodulin precurs |
| 21 | 104.5 | 3.3 | 644 | 2 I84634 | Tamm-Horsfall prot |
| 22 | 104.5 | 3.3 | 956 | 2 A57121 | thrombospondin 3 p |
| 23 | 100.5 | 3.1 | 1218 | 2 S71376 | glutamate receptor |
| 24 | 100 | 3.1 | 1993 | 2 T30902 | sodium channel SCA |
| 25 | 99 | 3.1 | 669 | 2 I38029 | matrix metalloprot |
| 26 | 98.5 | 3.1 | 459 | 2 JC5139 | vitronectin precur |
| 27 | 98 | 3.1 | 2531 | 2 T31070 | notch homolog - se |
| 28 | 98 | 3.1 | 2871 | 2 A55567 | fibrillin I - bovi |
| 29 | 97.5 | 3.0 | 365 | 2 A54785 | preadipocyte facto |

30 97.5 3.0 3002 2 A47221
31 97 3.0 862 2 H82182
32 97 3.0 2918 2 A54105
33 96.5 3.0 723 1 VCPVPP
34 96.5 3.0 986 2 G65116
35 96.5 3.0 1203 2 A49175
36 96.5 3.0 2871 2 A55624
37 95 3.0 468 2 T12725
38 95 3.0 2907 2 A57278
39 95 3.0 3871 2 T22812
40 94.5 3.0 956 1 A46016
41 94 2.9 381 1 OOPF2
42 94 2.9 612 2 A54282
43 94 2.9 915 2 A49874
44 93.5 2.9 385 2 S53718
45 93.5 2.9 621 1 KGB0H1

ALIGNMENTS

RESULT 1

T27628

hypothetical protein ZC506.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T27628

R:Harris, B.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z20395

A:Accession: T27628

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-999 <WIL>

A:Cross-references: EMBL:Z47073; PIDN:CAA87374.1; GSPDB:GN00028; CESP:ZC506.4

A:Experimental source: Clone ZC506

C:Genetics:

A:Gene: CESP:ZC506.4

A:Map position: X

A:Introns: 32/2; 65/3; 77/3; 113/2; 140/3; 177/2; 240/2; 277/1; 315/2; 439/2; 471/3;

C:Superfamily: metabotropic glutamate receptor 4

Query Match 4.3%; Score 136; DB 2; Length 999;

Best Local Similarity 22.3%; Pred. No. 0.027;

Matches 48; Conservative 44; Mismatches 87; Indels 36; Gaps 8;

QY 354 CKAGFYHGVLPVNNFRRGPDQHSIGTKDVSEAYV-----CLPCREGCPFCADDSPC 408

Db 615 CKIGFR-----KOLINDEQCCWACKCEDYEYLINETHCVGCGQWPTKDKGC 664

QY 409 F---VOEDKYLRL-----AITSFOGLCMLDFVSNLVVYHFRKAKSIRASGLILET 457

Db 665 FDLSELQKLYNWRSMYSLVPTILAVFGIATLVVVVYIYN--ETPVVAKSGRELSYI 722

QY 458 ILFGSLILLYFPVILYFEPSTFCILLRWALLGFARVYGVTVTLKHLRVLKVFLSRTAQR 517

Db 723 LLISIMCYCMTFVLLSKPSAIVCAIKRTGIGFAFCLYSAMFVKTNRIPIFSTRSAQR 782

QY 518 ---IPYMTGGVRMRLAVILLV--VFVFLI---GW 544

Db 783 PRFISPIQVVMVMTAMLAGVQLIGSLVLSVPPGW 817

RESULT 2

I59362

calcium/polyvalent cation-sensing receptor precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C:Accession: I59362; A55594

R:Ruot, M.; Molliver, M.E.; Snowman, A.M.; Snyder, S.H.

Proc. Natl. Acad. Sci. U.S.A. 92, 3161-3165, 1995

A:Title: Calcium sensing receptor: molecular cloning in rat and localization to nerve

C;Accession: A56715; S49341; A49419; B49419; C49419
J;Garrett, J.E.; Capuano, I.V.; Hamnerland, L.G.; Hung, B.C.P.; Brown, E.M.; Hebert,
R. Biol. Chem. 270, 12919-12925, 1995
A;Title: Molecular cloning and functional expression of human parathyroid calcium re
A;Reference number: A56715; MUID:95279439
A;Accession: A56715
A>Status: preliminary
A:Molecule type: mRNA
A;Residues: 1-1078 <GAR>
A;Cross-references: GB:I020759; NID:g683744; PIDN:AAA86503.1; PID:g683745
R;Pearce, S.H.S.; Thakker, R.V.
Submitted to the EMBL Data Library, August 1994
A;Reference number: S49341
A;Accession: S49341
A>Status: preliminary
A:Molecule type: DNA
A;Residues: 1-180, Q', 182-989, R', 991-1078 <PEA>
A;Cross-references: EMBL:X81086
R;Pollak, M.R.; Brown, E.M.; Chou, Y.H.; Hebert, S.C.; Marx, S.J.; Steinmann, B.; Lev
Cell 75, 1297-1303, 1993
A;Title: Mutations in the human Ca(2+)-sensing receptor gene cause familial hypocalc
A;Reference number: A49419; MUID:94094324
A;Accession: A49419
A>Status: preliminary
A:Molecule type: DNA
A;Residues: 289-303 <PO2>
A;Experimental source: family E
A;Note: sequence modified after extraction from NCBI backbone
A;Note: 186-Arg mutation is associated with familial hypocalciuric hypercalcemia and
A;Note: sequence extracted from NCBI backbone (NCBIN:142453)
A;Accession: B49419
A>Status: preliminary
A:Molecule type: DNA
A;Residues: 788-802 <PO3>
A;Experimental source: family J
A;Note: sequence modified after extraction from NCBI backbone
A;Note: 796-trp mutation is associated with familial hypocalciuric hypercalcemia and
A;Note: sequence extracted from NCBI backbone (NCBIN:142457)
C;Keywords: glycoprotein; receptor; transmembrane protein

Query Match 4.0%; Score 127; DB 2; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.15;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 OKYDIDOCSSDGWFSGTHKC---HLNNSE-CMPIKGLGVFLGAYEICAKAG---FYHGGV 363
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
475 EQTFDEC---GDVGNYSTIINHLLSPEDGSIVFKEVGI-----YNVYAKKGRLFTNEBK 527
QY : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 364 LPVNFRRRPGQH-----ISGSTKDVSSE-----AYVCLPCREG-----CPFCAD 404
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
528 ILWSGFREVFPFNCSRDCLAGTRKGIIEGEPTCCFECEVCEDPGEYSDETDASACNKCPD 587
QY : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 405 D-----SFCFFVDKYLR-----LAISFGCLCMLLDFVSMVLVVYHFRAKSTRASG 451
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
588 DFWSNENHTSCIAKETEFLSWTPEPFPGIALTLFAVLGIFLTAFVLGVFIKFRNTPIVKATN 647
QY : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

QY 452 LIILETILFSLLYFPVVTLYEPSTFRCILLRWARLLGFATVGTVTKLHRVLKVFL 511
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
548 RELSYLLLSLLCCFSSSLFFICEPDQWTCRLRPQAFGISFVLCISCIILVKNRVLLVFE 707
QY : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
512 SRTAQRIPYMTGGVRVMRLAV-----ILLVVWFELIGWTSV-CONLEKQISLI----- 559
Db : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
708 AKTIPSHRKWGLNLQFLVLCETMQVICWILYTAPPSRYRNROELEDIEFITCHE 767
QY : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

| | | | |
|----|-----|--|-----|
| Db | 648 | RELSYLLLFLLCCFSSUFFICEPDWTCRLRQPAFGISFVLCSILVKTNRVLVPE | 707 |
| QY | 512 | SRTAQRIPTYMTGGRVMRLAV-----ILLVVFELIGWTSV-CONLEKQISLI---- | 559 |
| Db | 708 | AKIPTSFRHKWGLNLOFLVFLCTMTOVICVILWLYTAPPSYRNOELEDIFITTCHE | 767 |

Db 538 LTFVLVQLQ-----VPFSNCSR-----DCLAGTRKGIIEGPTCCPECVCPDQ 581

QY 399 -----CPFCADD-----SPCFVQEDKYL-----LAIISFOGLCMLLDFVSM 435

Db 582 EYSDETDASACNKPDPFWSNHNHTSCIAKEIEFLSWTEPFGIALTLFVAGIFLTAFLV 641

QY 436 LVYHFRKAKSTRASGLILETLFGSLLLYPVWILYFEPSTFRICILLRWARKLLGFAIV 495

Db 642 GVPIKFRNTPIVKATNRELSYLLLFSLCCFSSSLFFIGEPQDWTCLRQPAFGISFVLC 701

QY 496 YGVTVTLKHLRVLFVLSRTAQRIPYMTGGRVVRMLAV-----ILLVFWELIGHTSSV 548

Db 702 ISCILVTKNRVLLVFAKIPTSFHRKMWGLNQLQFLVFLCTFMQIVICVIMLYTAPPSSY 761

QY 549 -CONLEKQISLI-----GQKTSDHILFNCLLI 575

762 RNOELEDIEIFITCHGSLMALGFLIGYTCLL 793

RESULT 7

JH0561

metabotropic glutamate receptor 2 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 13-Sep-1998

C:Accession: JH0561

R:Hanabe, Y.; Masu, M.; Ishii, T.; Shigemoto, R.; Nakanishi, S.

A:Title: A family of metabotropic glutamate receptors.

A:Reference number: JH0561; MUID:92110002

A:Accession: JH0561

A:Molecule type: mRNA

A:Residues: 1-872 <TAN>

A:Experimental source: brain

C:Comment: This protein is coupled to a G protein and evokes a variety of functions by m

C:Superfamily: metabotropic glutamate receptor 4

C:Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane prot

F:1-18/Domain: signal sequence #status predicted <Sig>

F:19-872/Product: metabotropic glutamate receptor 2 #status predicted <MET>

F:568-590/Domain: transmembrane #status predicted <TRI>

F:605-625/Domain: transmembrane #status predicted <TRII>

F:637-655/Domain: transmembrane #status predicted <TRIII>

F:680-700/Domain: transmembrane #status predicted <TRIV>

F:726-747/Domain: transmembrane #status predicted <TRV>

F:761-782/Domain: transmembrane #status predicted <TRVI>

F:795-819/Domain: transmembrane #status predicted <TRVII>

F:820-837/Domain: transmembrane #status predicted <TRVIII>

F:832/Binding site: phosphate (Ser) (covalent) #status predicted

F:832/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 3.6%; Score 116; DB 2; Length 872;

Best Local Similarity 18.9%; Pred. No. 0.79;

Matches 93; Conservative 63; Mismatches 184; Indels 152; Gaps 22;

QY 141 QSNKRE-----ONLQDDLDWTQALWSLLE-----GEPISRAAIFTFSDLSAPAPQVF 191

Db 271 REEDARELLAATQRLNASTTWASDGGWGALESVAVAGSERAAEGAITIELASPIISDFASY 330

QY 192 LOATREESRILLQDLSSAPHLANATLETWFHGLRRKWRPHILHRRGNQPGRLGHWSR 251

Db 331 FOS-----LDPWNN-----RNWP-----REFWEERFH-----CSFR 358

QY 252 RKDLGGDKSHKFPYPLECENGSKYKPGWLVTLSAIYGLQPNLVP----- 298

Db 359 ORD-----CAASHLRAVPFQESKIMFVNAVYMAHALNHMRALCPNTHLCDAMRPVN 414

QY 299 -----EPFGVMKVNDINLOKVDI-DQCSSDGFSGTHKRLANSECMPIKGLGVLGAYE 351

Db 415 GRRLYKDFVLNFKDAPFPADTDDEVRDFREGDG-----IGRN 454

QY 352 CIC-----KAGRYHPGVLPVNN-----RRGPDQHSIGSTKDYSE 387

Db 455 IFTYLRAGSGRYRYQKVGWAGGLDTSFIPWASPSAGPLPASRCEPCLQNEKVSQOP 514

QY 388 E---AYVCLPCREGCPF-----CADD-----SPCFVQEDKYL----- 417

Db 515 GEVCCWLCIPCO---PYEYRLDEFTCADGGLGWPNASLTGCFELPQEVIRWGDANAVGP 571

QY 418 LAIISFOGLCMLLDFVSMILVYHFRKAKSIRASGLILLLETLFGSLLLYPVWILYFEPSTFR 477

Db 572 VTACIGALATL---FVLGVFVRH-NATPVVKASGRGLCYILLGGVFLCYCMTFVFIKAPS 628

QY 478 TFRICILLRWARKLLGFAIVYMTGGRVVRMLAV-----LSRTAQRIPYMT-GGRVVRMLAVI- 533

Db 629 TAVCTLRRLGLGTAFSVCSYALLTNTNRIRIIFGARGAQRPRTFSPASQVAICLALIS 688

QY 534 ---LLVFWFLI 542

Db 689 GOLLIVAAWLVV 700

RESULT 8

A41939

G protein-coupled glutamate receptor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A41939; SI5362

R:Hanabe, Y.; Masu, M.; Ishii, T.; Shigemoto, R.; Nakanishi, S.

A:Title: Cloning, expression, and gene structure of a G protein-coupled glutamate rec

A:Reference number: A41939; MUID:92022526

A:Accession: A41939

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1199 <HOU>

A:Cross-references: GB:M61099; NID:g397806; PIDN:AAA19497.1; PID:g204460

A:Experimental source: cerebellum

A>Note: sequence extracted from NCBI backbone (NCBIP:60785)

R:Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi, S.

Nature 349, 760-765, 1991

A:Title: Sequence and expression of a metabotropic glutamate receptor.

A:Reference number: SI5362; MUID:91156047

A:Accession: SI5362

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1199 <MAS>

A:Cross-references: EMBL:X57569; NID:g56646; PIDN:CAA40799.1; PID:g56647

C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 3.6%; Score 115; DB 2; Length 1199;

Best Local Similarity 20.7%; Pred. No. 1.4;

Matches 61; Conservative 45; Mismatches 108; Indels 80; Gaps 9;

QY 357 GFYHPGVLPVNNFRRRPPDOHISGSKDYSE-----AYVCLPCREG- 398

Db 498 GTWHEGLVMDIDYKIQ---MNSGMVRSVCSEPCCLKGQIKVKGEVSCWICTACKNE 554

QY 399 -----CPFC-----ADDSPCFVQEDKYL-----LAIISFOGLCMLLDFVSMVLV 438

Db 555 FYODEFTCRACDLGWWPNAELTGCEPIPVRYLEWSDSIESIIAIAFSCIGILVTLFVLIF 614

QY 439 YHFRKAKSTRASGLILLLETLFGSLLLYPVWILYFEPSTFRICILLRWARKLLGFAIVYMT 498

Db 615 VLYRDTVPVVKSSRELICYIILAGIFLVGVCPTTLIAKPTTTCYLQRLVLGSSAMCYSA 674

QY 499 VTLKHLRVLV-----FLSRTAQRIPYMTGGRVVRMLAVIILVFWFLI----- 544

Db 675 LVYTKNTRTARILAGSKKIKCTKRPFRMSAWAOVIATISILISVQLTLVTLII----- 726

QY 545 TSSVCONLEKQISLQCKTSD-HLIFNNCLIDRWDMYTAGVMSLVSYDGLTI 597

Db 727 -----MEPPMILSYPSIKEVYLICN-----TSNLGWVAPVGVNGLLI 764

RESULT 9

[illegible]

A:Molecule type: protein

A:Residues: 128-162 <HI2>

A:Experimental source: PC-12 cell

C:Comment: This protein is a member of the epidermal growth factor family. It is functional in the differentiation of MDA-MB-453 cells.

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:361-397/Domain: EGF homology <EGF>

Query Match 3.4%; Score 109.5; DB 2; Length 868;
Best Local Similarity 19.1%; Pred. No. 2.5;
Matches 92; Conservative 57; Mismatches 173; Indels 159; Gaps 20;

```
Qy 31 PDSPRETPKPKPHQAQP-----GRASAS-----SSAPHSRSTGDTILA----- 71
Db 74 RPAAPPEPRPQPPQPSPAARRAAARSAAAGMRDRDPAGSGSMLLFGVSLACYSPSL 133
Qy 72 ---OKLAEPVMDVASYLY-----TGDShQLKRANCGRVELAGLCPKWPALASAPSL 122
Db 134 KSVQDOAYKAPVVVEGKVGQGLAPAGGSSNSTREPPASGRVALYKVLDKWP----- 184
Qy 123 HRALDTLTHATNFLNVLMSKNSREQNLQ----DDLWDYQALVMSLLEGEPSISRAAITF 178
Db 185 -----LRSGGLREQVIVSGSCAPLERNQRYIFEL---EP--TEQPLVF 223
Qy 179 STDLSAPAPQVFLQATREESRIILLODLSSSAPHLANATLETETWFGHLLRRKWRPHLHRRG 238
Db 224 KT--AFAPVDPNGKNIKKEVGKILCTDCAT-----RPLKKMK 259
Qy 239 PNOGPRGLGHSMRRKDGGLGGDKSHFKSPPYLECENGSYKPCWLVTLSAI---YG----- 291
Db 260 SOTGEVGKQSLKCAAGAGNQPQSRW-----FKDGKEINLRDRIRIKYGNRK 308
Qy 292 ---LQPNLVP-----EPRGMKVNDINLQKVDIDQCSS--DGWFSGTHKCH----- 331
Db 309 NSRLQFNKVKVEDAGEYVCEANILGKTVRGCHVNSVSTLSSWSGHARKCNETAKSY 368
Qy 332 -LNNECPKIGLGVLCAYECICKAGYHPCVLPVNNRRRGPD---OHISGSKTKVSE 387
Db 369 CVNGGVCIYIEG-----INQLSKCPNGFGQRCLEKPLRLYMPDPKQKHLGFELKE-AE 423
Qy 388 EAYVCLPCREGPCFCAADSPCFVQEDKYLRLAITSFOGLCMDFVSMVYVHPRKAKSI 447
Db 424 ELY-----QKRVTITICVALLVGVVAVCYCKTRKQ 457
Qy 448 R 448
Db 458 R 458
```

RESULT 15

3611

3611 protein - rat

A:Alternate names: angiotensinogen gene-inducible enhancer-binding protein; c-myc intron

C:Species: Rattus norvegicus (Norway rat)

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000

C:Accession: S53611; A39796; S22292; I58280

R:Makino, R.; Akiyama, K.; Yasuda, J.; Mashiyama, S.; Honda, S.; Sekiya, T.; Hayashi, K.

Nucleic Acids Res. 22, 5679-5685, 1994

A:Title: Cloning and characterization of a c-myc intron binding protein (MIBP1).

A:Reference number: S53611; MUID:95140632

A:Accession: S53611

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-2437 <NAK>

A:Cross-references: EMBL:D37951; NID:g1408559; PIDN:BAA07168.1; PID:g662296

R:Ron, D.; Brasier, A.R.; Habener, J.F.

Mol. Cell. Biol. 11, 2887-2895, 1991

A:Title: Angiotensinogen gene-inducible enhancer-binding protein 1, a member of a new fa

A:Reference number: A39796; MUID:91203912

A:Accession: A39796

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 'I', 1523-2263, 'K', 2265-2437 <RON>

A:Cross-references: GB:M65251; NID:g202790; PIDN:AAA00698.1; PID:g202791

R:Mitchellmore, C.; Traboni, C.; Cortese, R.

Nucleic Acids Res. 19, 141-147, 1991

A:Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha

A:Reference number: I58280; MUID:91187610

A:Accession: S22292

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 'E', 1729, 'R', 1734-2437 <MIT>

A:Cross-references: EMBL:X54249; NID:g57517; PIDN:CAA38150.1; PID:g57518

A:Note: the authors did not translate the codon for residue 1

C:Superfamily: HIV-EP2 enhancer-binding protein

C:Keywords: DNA binding; duplication; metal binding; transcription regulation; zinc f

F:191-211/Region: zinc finger CCHH motif

F:219-241/Region: zinc finger CCHH motif

F:934-940/Region: nuclear location signal

F:947-979/Region: serine-rich

F:1792-1812/Region: zinc finger CCHH motif

F:1820-1842/Region: zinc finger CCHH motif

F:1890-1917/Region: acidic

Query Match 3.4%; Score 107.5; DB 2; Length 2437;
Best Local Similarity 21.1%; Pred. No. 13;
Matches 71; Conservative 54; Mismatches 133; Indels 79; Gaps 15;

```
Qy 18 GLGAVGASRDQGRDPSRPTKPKHAQOPGRASADSSAPWSKSTDGTILAOKLAE 77
Db 1278 GASGLHKNLPKPFSDPGSKSTEAPTEQLLRDFASENACP-LQSLPGVVPVRIQTH 1336
Qy 78 VPMDEVASYLYTGDShQLKRANCGRYELAGLPGKWPALASAPHSILHRAIDTLTHATNFIN 137
Db 1337 VP-SYGSVMYTSISQILGONS-----PAIV-----ICKVDENMTQRTLVN 1376
Qy 138 VMLQS---NKREQNLQDDLDWYQALVMSLLEGE-----SISRAAITFST 180
Db 1377 AAMOGIGFNIAOVLGQRTGLEKYP--LWKVPQTLPLGLESSIPCLPSTDSAAISGSGK 1434
Qy 181 DSLAPAPQVFLQATREESRI-----LLQDLSSSAPHLANATLETETWFGHLLRRKWRP 232
Db 1435 RMLSPASSLELFEMETKQOKRVKEEKMYQIVVEL--SAVELTNSDIK-----KLSRPQKP 1488
Qy 233 HLHRRGPNQGRGLGHSWRRKDG-----GGDKSHFKWS---PPYLECENGSKYKPGWLVT 284
Db 1489 QLVROGCASEPKD-GSSQSRSSTSSLSFSSSQDHPAASGPPPPNPNRILSGSRAP----- 1542
Qy 295 LSSAIYGLQPNLVEPRGVMKVDINLQKVIDDQCSSD 321
Db 1543 -----PR--RKFGSPSESRESSDELDIDETSSD 1568
```

Search completed: November 1, 2001, 10:15:15

Job time: 813 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:20:45 ; Search time 24.77 Seconds
(without alignments)
828.383 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MCAMAYPLLLCLLLLAQLGLG.....YMTAVGMWSLVSVDGLTIFQ 599

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 161.5 | 5.0 | 976 | 1 MGR_DROME | P91685 drosophila |
| 2 | 136 | 4.3 | 999 | 1 MGR1_CAREL | Q09630 caenorhabdi |
| 3 | 134.5 | 4.2 | 1079 | 1 CASR_MOUSE | Q9qy96 mus musculu |
| 4 | 133.5 | 4.2 | 1079 | 1 CASR_RAT | P48442 rattus norv |
| 5 | 127 | 4.0 | 1078 | 1 CASR_HUMAN | P41180 homo sapien |
| 6 | 123 | 3.8 | 1085 | 1 CASR_BOVIN | P35384 bos taurus |
| 7 | 120 | 3.8 | 872 | 1 MGR2_HUMAN | Q14416 homo sapien |
| 8 | 116 | 3.6 | 872 | 1 MGR2_RAT | P31421 rattus norv |
| 9 | 116 | 3.6 | 1194 | 1 MGR1_HUMAN | Q13255 homo sapien |
| 10 | 115.5 | 3.6 | 877 | 1 MGR3_HUMAN | Q14832 homo sapien |
| 11 | 115 | 3.6 | 1199 | 1 MGR1_RAT | P23385 rattus norv |
| 12 | 114 | 3.6 | 1212 | 1 MGR5_HUMAN | P41594 homo sapien |
| 13 | 112 | 3.5 | 879 | 1 MGR3_RAT | P31422 rattus norv |
| 14 | 111 | 3.5 | 1203 | 1 MGR5_RAT | P31424 rattus norv |
| 15 | 109.5 | 3.4 | 868 | 1 MGR2_RAT | Q35569 rattus norv |
| 16 | 106.5 | 3.3 | 850 | 1 MGR2_HUMAN | Q14511 homo sapien |
| 17 | 106 | 3.3 | 383 | 1 DLK_HUMAN | P80370 homo sapien |
| 18 | 104.5 | 3.3 | 644 | 1 UROM_RAT | P27590 rattus norv |
| 19 | 104.5 | 3.3 | 956 | 1 TSP3_HUMAN | P49746 homo sapien |
| 20 | 101 | 3.2 | 912 | 1 MGR4_HUMAN | Q14833 homo sapien |
| 21 | 100 | 3.1 | 356 | 1 DCUP_DROME | Q9V595 drosophila |
| 22 | 99 | 3.1 | 669 | 1 MM15_HUMAN | P15111 homo sapien |
| 23 | 98.5 | 3.1 | 459 | 1 VTNC_PIG | P48819 sus scrofa |
| 24 | 98 | 3.1 | 2871 | 1 FBNI_BOVIN | P98133 bos taurus |
| 25 | 97.5 | 3.0 | 749 | 1 COA2_PAVPN | P22964 porcine par |
| 26 | 97.5 | 3.0 | 2871 | 1 FBNI_HUMAN | P35555 homo sapien |
| 27 | 97 | 3.0 | 915 | 1 MGR7_HUMAN | Q14831 homo sapien |
| 28 | 97 | 3.0 | 1078 | 1 SZ44_HUMAN | Q95486 homo sapien |
| 29 | 97 | 3.0 | 2911 | 1 FBNI_HUMAN | P35556 homo sapien |
| 30 | 96.5 | 3.0 | 986 | 1 YHDP_ECOLI | P46474 escherichia |
| 31 | 96.5 | 3.0 | 2871 | 1 FBNI_MOUSE | Q61554 mus musculu |
| 32 | 95 | 3.0 | 2907 | 1 FBNI_MOUSE | Q61555 mus musculu |
| 33 | 94.5 | 3.0 | 956 | 1 TSP3_MOUSE | Q05895 mus musculu |

ALIGNMENTS

RESULT 1

| ID | MGR_DROME | STANDARD; | PRT; | 976 AA. |
|----|--|-----------|------|---------|
| DT | P91685; | | | |
| DT | 15-JUL-1998 (Rel. 36, Created) | | | |
| DT | 15-JUL-1998 (Rel. 36, Last sequence update) | | | |
| DT | 15-JUL-1998 (Rel. 36, Last annotation update) | | | |
| DE | METABOTROPIC GLUTAMATE RECEPTOR PRECURSOR. | | | |
| GN | GLYRA OR GLU-RA. | | | |
| OS | Drosophila melanogaster (Fruit fly). | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | | |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | | |
| OX | NCBI_TaxID=7227; | | | |
| RN | [1]; | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=OREGON-R; | | | |
| RX | MEDLINE=96421661; PubMed=8824309; | | | |
| RA | Parmentier M.L., Pin J.P., Bockaert J., Grau Y.; | | | |
| RT | "Cloning and functional expression of a Drosophila metabotropic | | | |
| RT | glutamate receptor expressed in the embryonic CNS. "; | | | |
| RL | J. Neurosci. 16:6687-6694(1996). | | | |
| CC | -1- FUNCTION: RECEPTOR FOR GLUTAMATE. | | | |
| CC | -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. | | | |
| CC | -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE CNS OF THE LATE EMBRYO. | | | |
| CC | -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS. | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@sib-sib.ch). | | | |
| CC | EMBL: X99675; CAA67993.1; - | | | |
| DR | GDDB; GCR_1123; | | | |
| DR | FlyBase; FBgn0019985; Glu-RA. | | | |
| DR | InterPro; IPR000162; - | | | |
| DR | InterPro; IPR000337; - | | | |
| DR | InterPro; IPR001828; - | | | |
| DR | Pfam; PF00003; 7tm_3; 1. | | | |
| DR | Pfam; PF01094; ANF_receptor; 1. | | | |
| DR | PRINTS; PR00248; GPCRMR. | | | |
| DR | PRINTS; PR00593; MTABOTROPICR. | | | |
| DR | PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1. | | | |
| DR | PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1. | | | |
| DR | PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1. | | | |
| DR | PROSITE; PS02529; G_PROTEIN_RECEP_F3_4; 1. | | | |
| KW | G-protein coupled receptor; Transmembrane; Glycoprotein; Signal. | | | |
| FT | SIGNAL 1 ? | | | |
| FT | CHAIN ? 976 | | | |
| FT | DOMAIN ? 626 | | | |
| FT | DOMAIN 627 649 | | | |
| FT | TRANSMEM 650 663 | | | |
| FT | DOMAIN 664 684 | | | |
| FT | TRANSMEM 664 684 | | | |
| FT | II (POTENTIAL). | | | |
| FT | II (POTENTIAL). | | | |

| | | | | | | |
|----|------|-----|------|---|------------|--------------------|
| 34 | 94 | 2.9 | 381 | 1 | OPS2_DROME | P08099 drosophila |
| 35 | 94 | 2.9 | 915 | 1 | MGR7_RAT | P35400 rattus norv |
| 36 | 94 | 2.9 | 1403 | 1 | NID2_MOUSE | O88322 mus musculu |
| 37 | 93.5 | 2.9 | 385 | 1 | DLK_MOUSE | Q09163 mus musculu |
| 38 | 93.5 | 2.9 | 621 | 1 | KNHL_BOVIN | P01044 bos taurus |
| 39 | 93.5 | 2.9 | 808 | 1 | FTFB_DROME | O05192 drosophila |
| 40 | 93.5 | 2.9 | 840 | 1 | SYL_BORBU | O31267 borrelia bu |
| 41 | 93.5 | 2.9 | 1133 | 1 | EGF_RAT | P07522 rattus norv |
| 42 | 93 | 2.9 | 810 | 1 | NEL1_HUMAN | Q92832 homo sapien |
| 43 | 93 | 2.9 | 908 | 1 | MGR8_MOUSE | P47743 mus musculu |
| 44 | 92.5 | 2.9 | 810 | 1 | NEL1_RAT | O62919 rattus norv |
| 45 | 92.5 | 2.9 | 1184 | 1 | FBL2_HUMAN | P98095 homo sapien |

01-OCT-2000 (Rel. 40, Last annotation update)
EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
CELL CALCIUM-SENSING RECEPTOR).
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
STRAIN=C57BL/6; TISSUE=Kidney;
MEDLINE=20092890; PubMed=10625662;
Oda Y., Tu C.-L., Chang W., Crumrine D., Koemueves L., Mauro T.,
Elias P.M., Bikie D.D.;
RA "The calcium sensing receptor and its alternatively spliced form in
murine epidermal differentiation.";
J. Biol. Chem. 275:1183-1190(2000).
[2]
SEQUENCE FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
STRAIN=BLAK SWISS X 129/SVJ; TISSUE=Kidney;
MEDLINE=20119279; PubMed=10652312;
Pi M., Garner S.C., Flannery P., Spurney R.F., Quarles L.D.;
RA "Sensing of extracellular cations in CasR-deficient osteoblasts.
Evidence for a novel cation-sensing mechanism.";
J. Biol. Chem. 275:3256-3263(2000).
[3]
SEQUENCE OF 256-600 FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
TISSUE=Epiphyseal cartilage;
MEDLINE=20043955; PubMed=10579354;
Chang W., Tu C., Chen T.-H., Koemueves L., Oda Y., Pratt S.A.,
Miller S., Shoback D.;
RA "Expression and signal transduction of calcium-sensing receptors in
cartilage and bone.";
Endocrinology 140:5883-5893(1999).
[4]
SEQUENCE OF 507-582 FROM N.A. (ISOFORM A).
STRAIN=NMRI; TISSUE=Brain;
Hildenbrand J., Ammon H.P.T., Wahl M.A.;
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE OF 562-814 FROM N.A.
TISSUE=Kidney;
Moawad T.I., Riccardi D.;
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[6]
SEQUENCE OF 646-799 FROM N.A., AND TISSUE SPECIFICITY.
MEDLINE=97231187; PubMed=9076582;
Quarles L.D., Hartle J.E. II, Siddhanti S.R., Guo R., Hinson T.K.;
RA "A distinct cation-sensing mechanism in MC3T3-E1 osteoblasts
functionally related to the calcium receptor.";
J. Bone Miner. Res. 12:393-402(1997).
CC -!- FUNCTION: SENSES CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, KIDNEY AND CARTILAGE.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; AF110178; AAD28371.1; -;
EMBL; AF110179; AAD28372.1; -;
EMBL; AF128842; AAD40638.1; -;
EMBL; AF068900; AAC19388.1; -;

DR EMBL; AB027140; BAA77688.1; -;
DR EMBL; AF002015; AAC53252.1; -;
DR EMBL; AF159565; AAF00193.1; -;
DR MGD; MGI:1351351; Gprc2a.
DR InterPro; IPR000088; -;
DR InterPro; IPR000337; -;
DR InterPro; IPR001828; -;
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00592; CASSENSINGR.
DR PROSITE; PS00979; G_PROTEIN_RECP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
DR PROSITE; PS02259; G_PROTEIN_RECP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 19 POTENTIAL
FT CHAIN 20 1079 EXTRACELLULAR CALCIUM-SENSING RECEPTOR.
FT DOMAIN 20 612 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 613 635 I (POTENTIAL).
FT DOMAIN 636 649 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 650 670 II (POTENTIAL).
FT DOMAIN 671 681 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 682 700 III (POTENTIAL).
FT DOMAIN 701 724 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 725 745 IV (POTENTIAL).
FT DOMAIN 746 762 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 770 792 V (POTENTIAL).
FT DOMAIN 793 805 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 806 828 VI (POTENTIAL).
FT DOMAIN 829 836 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 837 862 VII (POTENTIAL).
FT DOMAIN 863 1079 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 594 594 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPLIC 461 537 MISSING (IN ISOFORM B).
FT CONFLICT 45 45 A -> S (IN REF. 2).
FT CONFLICT 304 304 L -> P (IN REF. 3).
FT CONFLICT 410 410 G -> D (IN REF. 2 AND 3).
FT CONFLICT 566 566 V -> A (IN REF. 2).
FT CONFLICT 595 595 Y -> H (IN REF. 2, 3 AND 5).
FT CONFLICT 610 610 E -> V (IN REF. 5).
FT CONFLICT 814 814 F -> L (IN REF. 5).
FT CONFLICT 889 889 L -> I (IN REF. 2).
FT CONFLICT 906 909 TGSN -> SGWI (IN REF. 2).
FT CONFLICT 1057 1057 V -> M (IN REF. 2).
FT CONFLICT 1064 1064 V -> A (IN REF. 2).
FT CONFLICT 1076 1076 I -> V (IN REF. 2).
SQ SEQUENCE 1079 AA; 120839 MW; AAF8D8D472736D6E CRC64;

Query Match 4.2%; Score 134.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.012;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGNPGRLGHWSRRKDLGGDKSHFKWSPYBCEGNSYKPGWLVTLSAIYGLQP 294
Db 377. HEEG---GNRLNLSSTAFRLPCTGDENINSVETPYMGVEHLIRISYNYLVAYISTAHALQD 433
QY 295 --NLVPFRGVMKVDINLQKVDIDQCSDDGFWGTHKCHLNNS-----ECMPTKG 342
Db 434 IYTCPLG-RGLF---TNGSCADIKV--EAQVVKLHLNFTNNMGSEQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRGPDQHS----- 379

```
Db 488 NYSIIINHLSPEDGSIIVKEGVYINVYAKKGERLFINEGKILWSGFSREVFPSCSRDQ 547
QY 380 -GSTKDVSE----AVVCLPREG-----CPFCADD-----SPCFVQEDKYL 417
Db 548 AGTRKGIIEGPTCCFECVCEPDGEYSGTDSACDKCPDDFHSNENYTSIAKEIEFLA 607
QY 418 -----LAISFQGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471
Db 608 WTEPFGIALTFLAVLGIFLTAFLGVFIKFRNTPIVKATNRELSYLLLSLCCFSSSLF 667
QY 472 LYEPSTFCILLRWALLGFAIVYGVTVLKLHRLVKVFLSRTAQRIPIYMTGGRVWRMLA 531
Db 668 FIEGPDQWTCRLRQAPAGISFVLCISILVKTNRVLVFEAKIPTSFHRKMWGLNLQFL 727
QY 532 V-----ILLVFWFLIGWTSSVCNLEKQISLIGOGKTSOHLIFNMCLIDRWDMYTA 584
Db 728 VFICTFMQIICIIWLYTAPPSSY-RNHELE-----DELIITC---HEGSLMAL 773
QY 585 GMSLVSY 592
Db 774 G--SLIGY 779

RESULT 4
CASR_RAT
ID CASR_RAT STANDARD; PRT; 1079 AA.
AC P48442;
DC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
DE CELL CALCIUM-SENSING RECEPTOR).
GN CASR OR PCAR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney outer medulla;
RX MEDLINE=95116508; PubMed=7816802;
RA Riccardi D., Park J., Lee W., Gamba G., Brown E.M., Hebert S.C.;
RT "Cloning and functional expression of a rat kidney extracellular
RT calcium/polyvalent cation-sensing receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:131-135(1995).
RN [2]
RP SEQUENCE OF 1-294 FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=95241465; PubMed=7724534;
RA Ruat M., Showman A.M., Snyder S.H.;
RT "Calcium sensing receptor: molecular cloning in rat and localization
RT to nerve terminals.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3161-3165(1995).
CC -!- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U10354; AAC52149.1; -.
CC EMBL; U20289; AAC52195.1; -.
CC GCRDb; GCR.1449; -.
Db InterPro; IPR000068; -.

```

```
DR InterPro; IPR000337; -.
DR InterPro; IPR001828; -.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00592; CASSENSINGR.
DR PROSITE; PS00979; G-PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G-PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G-PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS00981; G-PROTEIN_RECEP_F3_4; 1.
DR PROSITE; PS00981; G-PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1079
FT DOMAIN 20 612
FT TRANSMEM 613 635
FT DOMAIN 636 649
FT TRANSMEM 650 670
FT DOMAIN 671 681
FT TRANSMEM 682 700
FT DOMAIN 701 724
FT TRANSMEM 725 745
FT DOMAIN 746 769
FT TRANSMEM 770 792
FT TRANSMEM 793 805
FT TRANSMEM 806 828
FT TRANSMEM 829 836
FT TRANSMEM 837 862
FT TRANSMEM 863 1079
FT CARBOHYD 90 90
FT CARBOHYD 130 130
FT CARBOHYD 261 261
FT CARBOHYD 287 287
FT CARBOHYD 386 386
FT CARBOHYD 446 446
FT CARBOHYD 468 468
FT CARBOHYD 488 488
FT CARBOHYD 541 541
FT CARBOHYD 594 594
FT CARBOHYD 1079 AA; 120867 MW; D7664550361F9736 CRC64;
SQ SEQUENCE 1079 AA; 120867 MW; D7664550361F9736 CRC64;

Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.014;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGPNQGRGLGHWRKDGKSHKFPKSPVLECEGNSYKPGWLVTLSAIVGLQP 294
Db 377 HEEG---GNRLNSSTAFRPLCTGDNINSVETPTMDYEHRLISYNYLAVISIAHALQD 433
QY 295 --NLVPEFRGMKVDINLQKVDIQDCSDGWFSGTHKCHLNS-----ECMPIKG 342
Db 434 IYTCPLPG-RGLF--TNGSCADIKKV--EAWQVLKHLRLHFTNNMGEQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRGPDQHS---- 379
Db 488 NYSIIINHLSPEDGSIIVKEGVYINVYAKKGERLFINEGKILWSGFSREVFPSCSRDQ 547
QY 380 -GSTKDVSE----AVVCLPREG-----CPFCADD-----SPCFVQEDKYL 417
Db 548 AGTRKGIIEGPTCCFECVCEPDGEYSGTDSACDKCPDDFHSNENYTSIAKEIEFLA 607
QY 418 -----LAISFQGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471
Db 608 WTEPFGIALTFLAVLGIFLTAFLGVFIKFRNTPIVKATNRELSYLLLSLCCFSSSLF 667
QY 472 LYEPSTFCILLRWALLGFAIVYGVTVLKLHRLVKVFLSRTAQRIPIYMTGGRVWRMLA 531
Db 668 FIEGPDQWTCRLRQAPAGISFVLCISILVKTNRVLVFEAKIPTSFHRKMWGLNLQFL 727
QY 532 V-----ILLVFWFLIGWTSSVCNLEKQISLIGOGKTSOHLIFNMCLIDRWDMYTA 584
Db 728 VFICTFMQIICIIWLYTAPPSSY-RNHELE-----DELIITC---HEGSLMAL 773

```


Qy 585 GWMSLVSY 592
Db 774 G--SLIGY 779

RESULT 5
ID CASR_HUMAN STANDARD; PRT: 1078 AA.
AC P41180; Q13912; Q16379; Q16108; Q16109; Q16110;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
DE CELL CALCIUM-SENSING RECEPTOR).
GN CASR OR PCAR1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RA Pearce S.H.S., Thakker R.V.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[2]
RA SEQUENCE FROM N.A.
RA TISSUE=Parathyroid;
RA MEDLINE=95279439; PubMed=7759551;
RA Garrett J.E., Capuano I.V., Hammerland L.G., Hung B.C., Brown E.M.,
RA Hebert S.C., Nemeth E.F., Fuller F.;
RA "Molecular cloning and functional expression of human parathyroid
RA calcium receptor cDNAs";
RA J. Biol. Chem. 270:12919-12925(1995).
[3]
RA SEQUENCE FROM N.A.
RA TISSUE=Kidney;
RA MEDLINE=95408281; PubMed=7677761;
RA Aida K., Koishi S., Tawata M., Onaya T.;
RA "Molecular cloning of a putative Ca(2+)-sensing receptor cDNA from
RA human kidney";
RA Biochem. Biophys. Res. Commun. 214:524-529(1995).
[4]
RA SEQUENCE FROM N.A.
RA MEDLINE=96343808; PubMed=8756555;
RA Freichel M., Zink-Lorenz A., Hollloschi A., Hafner M., Flockerzi V.,
RA Raue F.;
RA "Expression of a calcium-sensing receptor in a human medullary
RA thyroid carcinoma cell line and its contribution to calcitonin
RA secretion";
RA Endocrinology 137:3842-3848(1996).
[5]
RA VARIANTS FHH GLU-185; LYS-297 AND TRP-795.
RA MEDLINE=94094324; PubMed=7916660;
RA Pollak M.R., Brown E.M., Chou Y.H., Hebert S.C., Marx S.J.,
RA Steinmann B., Levi T., Seidman C.E., Seidman J.G.;
RA "Mutations in the human Ca(2+)-sensing receptor gene cause familial
RA hypocalcemic hypercalcaemia and neonatal severe
RA hyperparathyroidism";
RA Cell 75:1297-1303(1993).
[6]
RA VARIANT ADH ALA-127.
RA MEDLINE=95179179; PubMed=7874174;
RA Pollak M.R., Brown E.M., Estep H.L., McLaine P.N., Kifor O., Park J.,
RA Hebert S.C., Seidman C.E., Seidman J.G.;
RA "Autosomal dominant hypocalcaemia caused by a Ca(2+)-sensing receptor
RA gene mutation";
RA Nat. Genet. 8:303-307(1994).
[7]
RA VARIANTS FHH MET-62; CYS-66; MET-138; GLU-143 AND GLN-227.
RA MEDLINE=95243222; PubMed=7726161;
RA Chou Y.-H.W., Pollak M.R., Brandi M.L., Toss G., Arnqvist H.,
RA Atkinson A.B., Papapoulos S.E., Marx S., Brown E.M., Seidman J.G.,
RA Seidman C.E.;
RA "Mutations in the human Ca(2+)-sensing-receptor gene that cause

RT familial hypocalcemic hypercalcaemia";
RL Am. J. Hum. Genet. 56:1075-1079(1995).
[8]
RP SEQUENCE OF 1-61 FROM N.A., AND VARIANT FHH ALA-39.
RX MEDLINE=95403641; PubMed=7673400;
RA Aida K., Koishi S., Inoue M., Nakazato M., Tawata M., Onaya T.;
RT "Familial hypocalcemic hypercalcaemia associated with mutation in the
RT human Ca(2+)-sensing receptor gene";
J. Clin. Endocrinol. Metab. 80:2594-2598(1995).
[9]
RN VARIANTS NSHPT LEU-227 AND TYR-598.
RP MEDLINE=96292293; PubMed=8675635;
RX Pearce S.H.S., Trump D., Wooding C., Besser G.M., Chew S.L.,
RA Grant D.B., Heath D.A., Hughes I.A., Paterson C.R., Whyte M.P.,
RA Thakker R.V.;
RT "Calcium-sensing receptor mutations in familial benign hypercalcaemia
RT and neonatal hyperparathyroidism";
J. Clin. Invest. 96:2683-2692(1995).
[10]
RN VARIANTS ADHP THR-116; HIS-681 AND SER-806, AND VARIANT SER-851.
RP MEDLINE=96311554; PubMed=8733126;
RX Baron J., Winer K.K., Yancovski J.A., Cunningham A.W., Laue L.,
RA Zimmerman D., Cutler G.B. Jr.;
RT "Mutations in the Ca(2+)-sensing receptor gene cause autosomal
RT dominant and sporadic hypoparathyroidism";
Hum. Mol. Genet. 5:601-606(1996).
[11]
RN VARIANT FHH ARG-174.
RP MEDLINE=97442275; PubMed=9298824;
RX Ward B.K., Stuckey B.G.A., Gutteridge D.H., Laing N.G., Pullan P.T.,
RA Ratajczak T.;
RT "A novel mutation (L174R) in the Ca2+-sensing receptor gene
RT associated with familial hypocalcemic hypercalcaemia";
Hum. Mutat. 10:233-235(1997).
CC -1- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS SEEM TO BE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: FOUND IN KIDNEY, BUT NOT IN BRAIN, LUNG,
CC LIVER, HEART, SKELETAL MUSCLE, OR PLACENTA.
CC -1- DISEASE: DEFECTS IN CASR ARE A CAUSE OF FAMILIAL HYPOCALCAEMIC
CC HYPERCALCAEMIA (FHH) AND NEONATAL SEVERE HYPERPARATHYROIDISM
CC (NSHPT). TWO INHERITED CONDITIONS CHARACTERIZED BY ALTERED CALCIUM
CC HOMEOSTASIS. THE MUTATIONS REDUCE THE ACTIVITY OF THE RECEPTOR.
CC FHH AFFECTED INDIVIDUALS EXHIBIT MILD OR MODEST HYPERCALCAEMIA,
CC RELATIVE HYPOCALCAEMIA, AND INAPPROPRIATELY NORMAL PTH LEVELS. IN
CC CONTRAST, NSHPT IS A RARE AUTOSOMAL RECESSIVE LIFE-THREATENING
CC DISORDER CHARACTERIZED BY VERY HIGH SERUM CALCIUM CONCENTRATIONS,
CC SKELETAL DEMINERALIZATION, AND PARATHYROID HYPERPLASIA. IN SOME
CC INSTANCES NSHPT HAS BEEN DEMONSTRATED TO BE THE HOMOZYGOUS FORM OF
CC FHH.
CC -1- DISEASE: DEFECTS IN CASR ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT
CC HYPOCALCAEMIA (ADH) IN WHICH THE RECEPTOR IS ACTIVATED AT SUBNORMAL
CC CA(2+) LEVELS.
CC -1- DISEASE: DEFECTS IN PCAR1 ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT
CC HYPOPARATHYROIDISM (ADHP). ADHP IS CHARACTERIZED BY HYPOCALCAEMIA
CC AND HYPERPHOSPHATEMIA DUE TO INADEQUATE SECRETION OF PARATHYROID
CC HORMONE. SYMPTOMS ARE SEIZURES, TETANY AND CRAMPS.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81086; CAA56990.1; -;
DR EMBL; U20759; AAA86503.1; -;

DR EMBL: U20760; AAA86504.1; -
DR EMBL: D50855; BAA09453.1; -
DR EMBL: S81176; AAB46873.1; -
DR EMBL: S79217; AAB35262.1; -
DR EMBL: S68032; AAB29413.2; ALT_SEQ.
DR EMBL: S68033; AAB29414.1; -
DR EMBL: S68036; AAB29415.1; -
DR GCRDB: GCR_1337; -
DR GCRDB: GCR_1874; -
DR GCRDB: GCR_2012; -
DR GCRDB: GCR_2013; -
DR GCRDB: GCR_2696; -
DR GCRDB: GCR_2697; -
DR MIM: 601199; -
DR MIM: 145980; -
DR MIM: 601198; -
DR InterPro: IPR000068; -
DR InterPro: IPR000337; -
DR InterPro: IPR001828; -
DR Pfam: PF01094; ANF_receptor; 1.
DR PRINTS: PR00248; GPCRMR.
DR PRINTS: PR00592; CASENSINGR.
DR PROSITE: PS00879; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE: PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
DR PROSITE: PS00259; G_PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Disease mutation; Alternative splicing; Polymorphism.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1078 EXTRACELLULAR CALCIUM-SENSING RECEPTOR.
FT DOMAIN 20 612 EXTRACELLULAR POTENTIAL.
FT TRANSMEM 613 635 I (POTENTIAL).
FT DOMAIN 636 649 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 650 670 II (POTENTIAL).
FT DOMAIN 671 681 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 682 700 III (POTENTIAL).
FT DOMAIN 701 724 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 725 745 IV (POTENTIAL).
FT DOMAIN 746 769 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 770 792 V (POTENTIAL).
FT DOMAIN 793 805 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 806 828 VI (POTENTIAL).
FT DOMAIN 829 836 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 837 862 VII (POTENTIAL).
FT DOMAIN 863 878 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 446 446 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 594 594 E -> EPLTFVLSVLQ (IN SECOND ISOFORM).
FT VARSPLIC 536 536 P -> A (IN FHH).
FT VARIANT 39 39 /FTIQ=VAR_003585.
FT VARIANT 52 62 R -> M (IN MILD FHH AND NSHPT).
Query Match 4.0%; Score 127; DB 1; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.045;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVDIDQSSDGFSTHCK---HLNNS-CMPKGLGVLCAYECICAKG---FVHPGV 363
DB 475 EQVTFDEC---GDLVGNYSIINHLNSPDGSIIVKEVGY---YNNYAKKGRFLFNEEK 527
QY 364 LPVNNFRRRGPDQH-----ISGSTKDVSSE---AVVCLUPCREG-----CPFCAD 404
DB 528 ILWSGFSRVPFNSRDLGTRKGIIEGPTCCFECVCEPDGYSDDTASACKNCPD 587

QY 405 D-----SPCFVQEDKYLR-----LAIISFQGLCMLLDVFSMLVYVYHFRKAKSIRASG 451
DB 588 DPWSNENHTSCIATKEITFLSWTEPFQIAULTFAVLGIELTAFLGVFIKFRNTPIVKATN 647
QY 452 LILLETILFGLSLLLYFPVIVILPEPSTFRFCILLRWALLGFAIVYGVTVTLKLRVLEKVL 511
DB 648 RELSYLLFLSLCCFSFSSFFIGEPQDWTCRLRQPAFGISFVLCISCLILVKTNRVLVEE 707
QY 512 SRTAQRIPYMTGGRVMRLAV-----ILLVFWFLIGWTSSV-CONLEKOISLI----- 559
DB 708 AKIPTSFRHKRWGLNLQFLVFLCTFMQIVICVIMLYTAPPSSYRNQOELEDEIIFITCHE 767
QY 560 GCGKTSDDLIFNMCLI 575
DB 768 GSLMALGFLIGYTCLL 783
RESULT 6
CASR_BOVIN STANDARD; PRT; 1085 AA.
AC P35384;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID
DE CELL CALCIUM-SENSING RECEPTOR).
GN CASR OR PCARL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Parathyroid; PubMed=8255296;
RX MEDLINE=94077182; PubMed=8255296;
RA Brown E.M., Gamba G., Riccardi D., Lombardi M., Butters R., Kifor O.,
RA Sun A., Hediger M.A., Lytton J., Hebert S.C.;
RT "Cloning and characterization of an extracellular Ca(2+)-sensing
RT receptor from bovine parathyroid.";
RL Nature 366:575-580(1993).
CC -!- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: S67307; AAB29171.1; -
CC PIR: S40476; S40476.
CC GCRDB: GCR_0900; -
CC InterPro: IPR000068; -
CC InterPro: IPR000337; -
CC InterPro: IPR001828; -
CC Pfam: PF01094; ANF_receptor; 1.
CC PRINTS: PR00248; GPCRMR.
CC PRINTS: PR00592; CASENSINGR.
CC PROSITE: PS00879; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE: PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE: PS00259; G_PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 19 POTENTIAL.

```
FT CHAIN 20 1085 EXTRACELLULAR CALCIUM-SENSING RECEPTOR.
FT DOMAIN 20 613 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 614 636 I (POTENTIAL).
FT DOMAIN 637 650 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 651 671 II (POTENTIAL).
FT DOMAIN 672 682 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 683 701 III (POTENTIAL).
FT DOMAIN 702 725 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 726 746 IV (POTENTIAL).
FT DOMAIN 747 770 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 771 793 V (POTENTIAL).
FT DOMAIN 794 806 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 807 829 VI (POTENTIAL).
FT DOMAIN 830 837 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 838 863 VII (POTENTIAL).
FT DOMAIN 864 1085 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 288 288 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 595 595 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1085 AA; 121170 MW; 5D66DE8C9CD13E47 CRC64;

Query Match 3.8%; Score 123; DB 1; Length 1085;
Best Local Similarity 21.18; Pred. No. 0.094;
Matches 64; Conservative 51; Mismatches 135; Indels 54; Gaps 12;

Qy 311 QKVIDQSSDGFSGTHK---HLNSE-CMPIKGLGVLGAYECICKAG---FYHPGV 363
Db 476 EQVTFDEC---GDLAGNYSIINWHLSPDGSIVKEVGY----YNYAKKGERLFINDEK 528

Qy 364 LPVNNFRFRGPDQH-----ISGSTKDVSE---AYVCLPREG-----CPFCAD 404
Db 529 ILWGSFREVFPSCNRCLAGTRKGIIEGTEPTCCFECVCPDGEYSDETDASACDKCPD 588

Qy 405 D-----SPCFVQEDKYLR-----LAIISFGICMLLDFVSMVYVHFRKAKSIASG 451
Db 589 DFWSNENHTSCIAKEIEFLSWTEPFGIALTFVLGIFLTFVGLFKFRNTPIVIRATN 648

Qy 452 LILLETILFGLSLLYFPVYLYFFSTFRCLLRWARLLGFATVYGVTLKLRVLYKVL 511
Db 649 RELSVLLFLSLCCFSSSLFFIGEPQDWTCRLRQPAFGISFVLCSILVKTNRVLLVFE 708

Qy 512 SRTAQRIPYMTGGRVMRLAVILLVWFVFLIG--WTSSVCNLEKQISLIGQGKTSDDL 569
Db 709 AKIPTSFRKWWGLNQLQFLVFLCTFMQIVICAIWLNATAPPSSYRNHEL-----EDBII 762

Qy 570 FNMIC 573
Db 763 FITC 766

RESULT 7
MGR2_HUMAN STANDARD; PRT; 872 AA.
ID MGR2_HUMAN
AC Q14416;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 2 PRECURSOR.
GN GRM2 OR MGLUR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=95346007; PubMed=7620613;
RA Fior P.J.; Lindauer K.; Puttner I.; Ruegg D.; Lukic S.; Knopfel T.;
RA Kuhn R.;
RT "Molecular cloning, functional expression and pharmacological
RT characterization of the human metabotropic glutamate receptor type
RL 2.";
RL Eur. J. Neurosci. 7:622-629(1995).
CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE ACTIVITY.
CC MAY MEDiate SUPPRESSION OF NEUROTRANSMISSION OR MAY BE INVOLVED IN
CC SYNAPTOGENESIS OR SYNAPTIC STABILIZATION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN DIFFERENT REGIONS OF THE
CC ADULT BRAIN AS WELL AS IN FETAL BRAIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR3.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35318; AAA76855.1;
CC GCRDB; GCR_1846;
CC MIM; 604099;
CC InterPro; IPR000162;
CC InterPro; IPR000337;
CC InterPro; IPR001458;
CC InterPro; IPR001828;
CC Pfam; PF00003; 7tm_3; 1.
CC Pfam; PF01094; ANF_receptor; 1.
CC PRINTS; PR00248; GPCRMR.
CC PRINTS; PR00593; METABOTROPICR.
CC PRINTS; PR01052; METABOTROPIC2R.
CC PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family.
KW SIGNAL
FT CHAIN 19 872 POTENTIAL.
FT DOMAIN 19 872 METABOTROPIC GLUTAMATE RECEPTOR 2.
FT TRANSMEM 568 590 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 591 604 I (POTENTIAL).
FT TRANSMEM 605 625 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 626 636 II (POTENTIAL).
FT TRANSMEM 637 655 III (POTENTIAL).
FT DOMAIN 656 679 IV (POTENTIAL).
FT TRANSMEM 680 700 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 701 725 V (POTENTIAL).
FT TRANSMEM 726 747 VI (POTENTIAL).
FT DOMAIN 748 760 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 761 783 VII (POTENTIAL).
FT DOMAIN 784 793 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 794 819 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DOMAIN 820 872 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 872 AA; 95507 MW; 058608C35C701E9D CRC64;
```

```
Query Match 3.8%; Score 120; DB 1; Length 872;
Best Local Similarity 26.28; Pred. No. 0.12;
Matches 49; Conservative 28; Mismatches 70; Indels 40; Gaps 9;

Qy 390 YVCLPFCRGCPFF-----CADD-----SPCFVQEDKYLR-----LAIS 422
```

[illegible]

01-OCT-2000 (Rel. 40, Last annotation update)
 METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.
 GRII OR MGLUR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96029774; PubMed=7476890;
 RA Desai M.A., Burnett J.P., Mayne N.G., Schoepp D.D.;
 RT "Cloning and expression of a human metabotropic glutamate receptor 1
 alpha: enhanced coupling on co-transfection with a glutamate
 transporter.";
 RT Mol. Pharmacol. 48:648-657(1995).
 RL [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=97231349; PubMed=9076744;
 RA Stephan D., Bon C., Holzwarth J.A., Galvan M., Pruss R.M.;
 RT "Human metabotropic glutamate receptor 1: mRNA distribution, splice
 variants.";
 RT Neuropharmacology 35:1649-1660(1996).
 RL [3]
 CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
 MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
 CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
 ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN
 THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
 CC STRONGEST, TO MGLUR5.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U31215; AAA87843.1; -
 CC EMBL; U31216; AAA87844.1; -
 CC EMBL; L76627; AAB05337.1; -
 CC EMBL; L76631; AAB05338.1; -
 CC MIM; 604473; -
 CC GCRDb; GCR_1825; -
 CC GCRDb; GCR_1826; -
 CC GCRDb; GCR_1982; -
 CC GCRDb; GCR_1983; -
 CC InterPro; IPR000162; -
 CC InterPro; IPR000337; -
 CC InterPro; IPR001256; -
 CC InterPro; IPR001828; -
 CC Pfam; PF00003; 7tm_3; 1.
 CC Pfam; PF01094; ANF_receptor; 1.
 CC PRINTS; PR00248; GPCRMR.
 CC PRINTS; PR00593; METABOTROPIC1.
 CC PRINTS; PR01051; MTABOTROPIC1.
 CC PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
 CC PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
 CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
 CC PROSITE; PS00982; G_PROTEIN_RECEP_F3_4; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 CC Multigene family; Alternative splicing.
 KW SIGNAL 1 18 POTENTIAL.
 KW CHAIN 19 1194 METABOTROPIC GLUTAMATE RECEPTOR 1.
 FT DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 593 615 I (POTENTIAL).
 FT DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 630 650 II (POTENTIAL).
 FT DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 662 680
 FT DOMAIN 681 706
 FT TRANSMEM 707 727
 FT DOMAIN 728 750
 FT TRANSMEM 751 772
 FT DOMAIN 773 785
 FT TRANSMEM 786 808
 FT DOMAIN 809 814
 FT TRANSMEM 815 840
 FT DOMAIN 841 1194
 FT DOMAIN 1014 1035
 FT DOMAIN 1067 1081
 FT DOMAIN 1095 1130
 FT DOMAIN 1142 1194
 FT CARBOHYD 98 98
 FT CARBOHYD 223 223
 FT CARBOHYD 397 397
 FT CARBOHYD 515 515
 FT VARSPLIC 887 906
 FT VARSPLIC 907 1194
 FT VARSPLIC 887 906
 FT CONFLICT 593 593
 FT SEQUENCE 1194 AA; 132376 MW; 970E51AF40584F40 CRC64;
 Query Match 3.6%; Score 116; DB 1; Length 1194;
 Best Local Similarity 22.1%; Pred. No. 0.38;
 Matches 49; Conservative 34; Mismatches 95; Indels 44; Gaps 6;
 QY 357 GFYHPGVLPVNNRRRGGPDQHSIGSTKDVSEE-----AVVCLPCRGG- 398
 Db 498 GTWHEGVNIDDDYKIQ---MKNKSGVRSVCSEPCLGKQIKVIRKEVSCCVCTACKENE 554
 QY 399 -----CPCP-----ADSPCFVQEDKYLK-----LAIISFGLCMLLDVSMVLV 438
 Db 555 YVQDEFTCKACDLGWPNADLTGCEPIPVRYLEWSNIEPIIAIAFSCILGILVTLFVTLIF 614
 QY 439 YHFKKASIRASGLILLETILFGSLLYFPVVIYFPPVILYFPPSTFCILLRWALLFATVYGT 498
 Db 615 VLYRDPVWKSRSRELCCYIILAGIFLGVCFPTLIARPTTSCYLLQRLVGLSSAMCYSA 674
 QY 499 VTLKHLRVKVL---SRTAQRIPYMTGGRVMRLAVILLV 537
 Db 675 LVTKNRIARILAGSKKIKCTKPKRFMSAMVAQVIAISILISV 716
 RESULT 10
 MGR3_HUMAN STANDARD; PRT; 877 AA.
 ID MGR3_HUMAN
 AC Q14832;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE METABOTROPIC GLUTAMATE RECEPTOR 3 PRECURSOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96437205; PubMed=8840013;
 RA Makoff A., Volpe F., Lechuk R., Harrington K., Emson P.;
 RT "Molecular characterization and localization of human metabotropic
 glutamate receptor type 3.";
 RL Brain Res. Mol. Brain Res. 40:55-63(1996).
 CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
 IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
 ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR2.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X77748; CAA54796.1; -
CC GCRDb: GCR_2070; -
CC MIM: 601115; -
CC InterPro: IPR000162; -
CC InterPro: IPR000337; -
CC InterPro: IPR001234; -
CC InterPro: IPR001828; -
CC Pfam: PF00003; 7tm_3; 1.
CC Pfam: PF01094; ANF_receptor; 1.
CC PRINTS: PRO0248; GPCRMR.
CC PRINTS: PRO0593; MTABOTROPICR.
CC PRINTS: PRO1053; MTABOTROPICR.
CC PROSITE: PS00979; G_PROTEIN_RECF_3_1; 1.
CC PROSITE: PS00980; G_PROTEIN_RECF_3_2; 1.
CC PROSITE: PS00981; G_PROTEIN_RECF_3_3; 1.
CC PROSITE: PS0259; G_PROTEIN_RECF_3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 877
FT DOMAIN 21 574
FT TRANSMEM 575 597
FT DOMAIN 598 611
FT TRANSMEM 612 632
FT DOMAIN 633 643
FT TRANSMEM 644 662
FT DOMAIN 663 686
FT TRANSMEM 687 707
FT DOMAIN 708 732
FT TRANSMEM 733 754
FT DOMAIN 755 767
FT TRANSMEM 768 790
FT DOMAIN 791 800
FT TRANSMEM 801 826
FT DOMAIN 827 877
FT CARBOHYD 207 207
FT CARBOHYD 290 290
FT CARBOHYD 412 412
FT CARBOHYD 437 437
FT SEQUENCE 877 AA; 98619 MW; 66F28663CE35F740 CRC64;
Query Match 3.6%; Score 115.5; DB 1; Length 877;
Best Local Similarity 18.6%; Pred. No. 0.27;
Matches 90; Conservative 72; Mismatches 196; Indels 125; Gaps 23;
QY 138 VMLQSKSRE-----ONLQDDLDWQALVW-----SLLEGPSISRAAIFSTDSLAPAP 188
DB 272 LFMRSDDSRELIATAASRANASTFWASDCGAQESIIGSEHVAYCAITLLEAS----- 325
QY 189 QVFLQNTRESRIILQDLSSAPHLANATLETWFHGLRRKWRPHLRGPNQGRGLGH 248
DB 326 ----QPVROFDR-YFQSLNPYNNH-----RNPWFRDE----- 352
QY 249 SWRRK--DGLGDKSHFKWSPPLCEGNSYKP-----GWLVTLSAIYGLQPNLVP 298
DB 353 -WEQFQCSLQNRHRRVCDRLHALIDSSNYDOESKIMFVNAVAVAMALHMKMORTLCP 411
QY 299 ----EPRGVKMDINLQKVDIQDCSSDGFWSGTHKCHLNNSCEMPKIGLGFVLGAYECI--- 353
DB 412 NTKLCDAWKI-LDGKKLYKDYLLKIN-FTAPFPNPKDADSVKFTDGDGNGRYNVNF 469

QY 354 -----CKAGFYHPGV-LPVNPF---RRGPDQHIS-----GSTKDVSEB---AYVC 392
DB 470 QNVGGKYSYLVKVGHWAEATLSLDVNSHWSRNSVPTSCQSDPCAPNEMKMQPDGVCCWIC 529
QY 393 LPCR-----GCPFC-----ADSPCFVQEDKYLK-----LAIISFOGL 426
DB 530 IPECEVEYLADEFTCMDCGQWPTADLTGCDLPEDYIRWEDAWAIGPVTTIACLGFMCT 589
QY 427 CMLLDFVSMVYVYHFRKAKSIRASGLILETILFGLSLLYFPVVIYLFEPSTFCILLRW 486
DB 590 CMV-----VTVEIKHNNTPLVRASGRELKYILLFGVGLSYCMFTFFIAKPSVICALRRL 644
QY 487 ARLLGPAFTVGTTLKLRVLFV---LSRTAQRIPTWT---GGRVMRLAVTL-----LVFVW 539
DB 645 GLGSFAICYALLTKTNCIARIFDGVKNGAQRKPFISPSQVFCIGLILVQIVMVSVM 704
QY 540 FLI 542
DB 705 LIL 707
RESULT 11
MGLR1_RAT
ID MGLR1_RAT STANDARD; PRT; 1199 AA.
AC P23385;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.
GN GRM1 OR MGLUR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=91156047; PubMed=1847995;
RA Masu M., Tanabe Y., Tsuchida K., Shigemoto R., Nakanishi S.;
RT "Sequence and expression of a metabotropic glutamate receptor.";
RL Nature 349:760-765(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9202526; PubMed=1656524;
RA Houamed K.M., Kujper J.L., Gilbert T.L., Haldeman B.A., O'Hara P.J.,
RA Mulvihill E.R., Almers W., Hagen F.S.;
RT "Cloning, expression, and gene structure of a G protein-coupled
glutamate receptor from rat brain.";
RL Science 252:1318-1321(1991).
RN [3]
RP ALTERNATIVE SPLICING (ISOFORM 1B).
RC TISSUE=Brain;
RX MEDLINE=92110002; PubMed=1309649;
RA Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT "A family of metabotropic glutamate receptors.";
RL Neuron 8:169-179(1992).
RN [4]
RP ALTERNATIVE SPLICING (ISOFORM 1C).
RC TISSUE=Brain;
RX MEDLINE=93066232; PubMed=1438218;
RA Pin J.-P., Waerber C., Prezeau L., Bockaert J., Heinemann S.F.;
RT "Alternative splicing generates metabotropic glutamate receptors
inducing different patterns of calcium release in Xenopus oocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10331-10335(1992).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM DEPRESSION IN
CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1A (SHOWN HERE), 1B AND 1C;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. 1B AND 1C ARE C-TERMINALLY

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```
CC EMBL; D28538; BAA05891.1; -
CC EMBL; D28539; BAA05892.1; -
CC EMBL; S64316; AAD13954.1; -
DR GCRDB; GCR_0761; -
DR GCRDB; GCR_1002; -
DR GCRDB; GCR_1003; -
DR GCRDB; GCR_1317; -
DR MIM; 604102; -
DR InterPro; IPR000162; -
DR InterPro; IPR000202; -
DR InterPro; IPR000337; -
DR InterPro; IPR001828; -
DR Pfam; PF00003; 7tm_3; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01085; MTABOTROPICR.
DR PROSITE; PS00979; G_PROTEIN_RECF_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECF_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECF_F3_3; 1.
DR PROSITE; PS00982; G_PROTEIN_RECF_F3_4; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 1212
FT DOMAIN 22 579
FT TRANSMEM 580 602
FT DOMAIN 603 616
FT TRANSMEM 617 637
FT DOMAIN 638 648
FT TRANSMEM 649 667
FT DOMAIN 668 693
FT TRANSMEM 694 714
FT DOMAIN 715 737
FT TRANSMEM 738 759
FT DOMAIN 760 772
FT TRANSMEM 773 795
FT DOMAIN 796 801
FT TRANSMEM 802 827
FT DOMAIN 828 1212
FT CARBOHYD 88
FT CARBOHYD 210
FT CARBOHYD 378
FT CARBOHYD 382
FT CARBOHYD 445
FT CARBOHYD 734
FT VARSPLIC 877
SQ SEQUENCE 1212 AA; 132468 MW; A3C73606681C6A25 CRC64;

Query Match
Best Local Similarity 3.6%; Score 114; DB 1; Length 1212;
Matches 80; Conservative 53; Mismatches 165; Indels 94; Gaps 16;

QY 219 ETWFHGLRRKWRPHLRGPNQGRGLGHSWR-----RKDGLGGDKSHFKWSPPLYECE 273
DB 333 DVKWFDDYLLKLRPETNHRNP-----WFQEFWQHRFQCRLEGFPQENSRYNKT-----CN 382

QY 274 NG-----SYKPGWLW-TLSSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311
DB 383 SSLTKTHVQSDKMGFVFNATYSMAYGLHNMQMSLCPGYAGLCDDAMKPIDGRKLLKSLM 442

QY 312 KYVIDQCSGDFGSGTHKHLNNSCMPKGLGLGVGLGAYECICKAGFYHPGVLPVNNFR 371
DB 443 KTNFTGVSGDTILFDENGDSRGYELMNFKMG-----KDYFDIYNGVSDNGEL 492

QY 372 RGPQDQHI-----SGSTKDVSEE-----AYVCLPCREG-----CPPC 402
DB 493 KMDDEVVSKSKNIIRSVSCPECKGQIKVRKGEVSCWCTCTPCKENEYVFDEYTCRAC 552
```

```
QY 403 -----ADD-SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIRA 449
DB 553 QLGSWPTDITLGTGDLIPVQYLRMGDPEPTAAVVFACGLGLATLFTVTVFIYRDTVPVK 612
QY 450 SGLILLETILFGLSLLYFPVILYFSPFRFCILLRWARLIGFATVYGTVTLKLHRLVKV 509
DB 613 SSRELCTIILAGICGLYLCFTCLIAKPKQIYCYLQIGIGLSPAMSYSALVTNTRIARI 672
QY 510 FLISRTAQRI-----PYMTGGRVMRMLAVILLV 537
DB 673 -LAGSKKICKTKKPREMSACAQLVIAFILICI 703

RESULT 13
MGR3_RAT
ID MGR3_RAT STANDARD; PRT; 879 AA.
AC P31422;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 3 PRECURSOR.
GN GRM3 OR MGLUR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92110002; PubMed=1309649;
RA Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT "A family of metabotropic glutamate receptors.";
RL Neuron 8:169-179(1992).
CC 1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC 1- TISSUE SPECIFICITY: IS WIDELY DISTRIBUTED IN THE CNS. PREDOMINANT
CC EXPRESSION IS SEEN IN THE NEURONAL CELLS OF THE CEREBRAL CORTEX,
CC DENTATE GYRUS, AND GLIAL CELLS THROUGHOUT BRAIN REGIONS.
CC 1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; M92076; -; NOT_ANNOTATED_CDS.
DR PIR; JH0562; JH0562.
DR GCRDB; GCR_03624.
DR InterPro; IPR000162; -.
DR InterPro; IPR000337; -.
DR InterPro; IPR001234; -.
DR InterPro; IPR001828; -.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01053; MTABOTROPICR.
DR PROSITE; PS00979; G_PROTEIN_RECF_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECF_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECF_F3_3; 1.
DR PROSITE; PS00982; G_PROTEIN_RECF_F3_4; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family.
FT SIGNAL 1 22
FT CHAIN 23 879
FT DOMAIN 23 576
FT TRANSMEM 577 599
FT I (POTENTIAL).
```



```
FT DOMAIN 600 613 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 614 634 II (POTENTIAL).
FT DOMAIN 635 645 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 646 664 III (POTENTIAL).
FT DOMAIN 665 688 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 689 709 IV (POTENTIAL).
FT DOMAIN 710 734 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 735 756 V (POTENTIAL).
FT DOMAIN 757 769 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 770 792 VI (POTENTIAL).
FT DOMAIN 793 802 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 803 828 VII (POTENTIAL).
FT DOMAIN 829 879 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 879 AA; 98959 MW; 3E5965EDD5E6DEED CRC64;

Query Match 3.5%; Score 112; DB 1; Length 879;
Best Local Similarity 22.6%; Pred. No. 0.52;
Matches 42; Conservative 30; Mismatches 76; Indels 38; Gaps 7;

Qy 390 YVCLPCRE-----GCPFC-----ADSPCFVQEDKYL-----LAISF 423
Db 529 WICIPCEYEVYVDEFTCMDCGPGQWPTADLSGCYNLPEDYIKWEDAWAIGVPTIACIGF 588
Qy 424 QGLCMLDFVSMVYHFRKAKSIRASGLILLETILFGLSLLYFPVIVLYEPSPFCIL 483
Db 589 LCTCIVI-----TVPIKHNTPLYKASGRCELYILLFGVLSYSCWTFEFAKPSVPICAL 643
Qy 484 LRWALLGFATYGVPTVTLKHLRVKVF--LSRTAQIPYMT--GGVVRMLAVIL----LV 536
Db 644 RRLGLTGFACYSALLTKTNCIARIFDGVKNQAPRPFISPSQVFCILGLILVQIVMV 703
Qy 537 VFWFLI 542
Db 704 SVWLIL 709

RESULT 14
MGR5_RAT
ID MGR5_RAT STANDARD; PRT; 1203 AA.
AC P31424;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.
GN GRM5 OR MGLUR5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=923117054; PubMed=1320017;
RA Abe T., Sugihara H., Nawa H., Shigemoto R., Mizuno N., Nakanishi S.;
RT "Molecular characterization of a novel metabotropic glutamate
RT receptor mGLUR5 coupled to inositol phosphate/Ca2+ signal
RT transduction.";
RL J. Biol. Chem. 267:13361-13368(1992).
RN [2]
RP SEQUENCE OF 859-923 FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=93343913; PubMed=7688218;
RA Minakami R., Katsuki F., Sugiyama H.;
RT "A variant of metabotropic glutamate receptor subtype 5: an
RT evolutionally conserved insertion with no termination codon.";
RL Biochem. Biophys. Res. Commun. 194:622-627(1993).
CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
```

```
CC CALCIIUM SECOND MESSENGER SYSTEM AND GENERATES A CALCIUM-ACTIVATED
CC CHLORIDE CURRENT.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 5A (SHOWN HERE) AND 5B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER BY AN INSERTION OF
CC 32 RESIDUES
CC -1- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN NEURONAL CELLS OF THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: ACTIVATED BY QUISQUALATE > GLUTAMATE > IBOTENATE >
CC TRANS-1- AMINOCYCLOPENTYL-1,3-DICARBOXYLATE.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10891; -; NOT_ANNOTATED_CDS.
CC EMBL; S64315; AAB27666.1; -;
CC PIR; A42916; A42916.
CC GCRdb; GCR_0444; -.
CC GCRdb; GCR_0760; -.
CC InterPro; IPR000162; -.
CC InterPro; IPR000202; -.
CC InterPro; IPR000337; -.
CC InterPro; IPR001828; -.
CC Pfam; PF00003; 7cml_3; 1.
CC Pfam; PF01094; ANF_receptor; 1.
CC PRINTS; PR00248; GPCRMR.
CC PRINTS; PR00593; MTABOTROPICR.
CC PRINTS; PR01055; MTABOTROPIC5R.
CC PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE; PS0259; G_PROTEIN_RECEP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family; Alternative splicing.
CC SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1203 METABOTROPIC GLUTAMATE RECEPTOR 5.
FT DOMAIN 22 578 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 579 601 I (POTENTIAL).
FT DOMAIN 602 615 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 616 636 II (POTENTIAL).
FT DOMAIN 637 647 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 648 666 III (POTENTIAL).
FT DOMAIN 667 692 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 693 713 IV (POTENTIAL).
FT DOMAIN 714 736 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 737 758 V (POTENTIAL).
FT DOMAIN 759 771 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 772 794 VI (POTENTIAL).
FT DOMAIN 795 800 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 801 826 VII (POTENTIAL).
FT DOMAIN 827 1203 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 876 907 MISSING (IN ISOFORM 5A).
SQ SEQUENCE 1203 AA; 131885 MW; 99CA51E9E1ICIEA4 CRC64;
Query Match 3.5%; Score 111; DB 1; Length 1203;
Best Local Similarity 20.8%; Pred. No. 0.95;
Matches 80; Conservative 54; Mismatches 171; Indels 80; Gaps 17;
```

Qy 219 ETEWPHGLRRKWRPHLRHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPYLECE 273

Db 332 DVKWDYYLKLRLPNTLRNP-----WQEFWQHRFOCRUEGFAQENSKYNKT-----CN 381
 Qy 274 NG-----SYKPGMLV-TLSSAIYGL---OPNLVPERGV-----MKVDINLQ 311
 Db 382 SSLTLRHHVQDSKMGFVINAIIYMAVGLHMQMSLCPGVAGLCDAMKPIDGRKLLDSLM 441
 Qy 312 KVIDQSSDQSGWSPGTHKHCHLNSECMPIKGLGVFLGAYECICAKGYPHGVLPV----- 366
 Db 442 KTNFTGVSGDMILFDENGDSFGRYEIMNFKEMG---KDYFDYINVGSWONGELKMDDEV 498
 Qy 367 -----NNFRRCPOHI-SGSTKDVSEE-----AYCLPCREG-----CPFC-----A 403
 Db 499 WSKNNIIRSVCEPCKGQIKVIRGEVSCCTCTPCKENYVDFEYTCACQLGSWPT 558
 Qy 404 DD-SPCFVQEDKYLK-----LAIISFQGLMLLDFVSMVLVYHFKAKSRASGLILLE 456
 Db 559 DDLTGCDLIPVQLRWGDPPIAAVFAVACGLGLLATLFTVIVIFIIYRDTPVVKSSRELVCY 618
 Qy 457 TILFGLLLFPVVIYFEPSTFRCILLRWALLGFATVGTVTLKLHRLVKVFLSRTAQ 516
 Db 619 IILAGICLGLYCTFCLIAKPKQYICLQIRIGLISLSPAMSYSLVTKTNRIARI-LAGSKK 677
 Qy 517 RI-----PYMTGGRRVRLAVILLV 537
 Db 678 KICTKPRFMSACQLVIAFILICI 702
 RESULT 15
 NR2_RAT
 ID NR2_RAT STANDARD; PRT: 868 AA.
 AC Q35569; Q35570; Q35571; Q35572; Q35073;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PRO-NEUREGULIN-2 PRECURSOR (PRO-NRG2) [CONTAINS: NEUREGULIN-2 (NRG-2)
 DE (NEURAL-AND THYMUS-DERIVED ACTIVATOR FOR ERBB KINASES) (NTAK)].
 GN NR2 OR NTAK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., SEQUENCE OF 128-162, AND ALTERNATIVE SPLICING.
 RX MEDLINE=98006324; PubMed=9348101;
 RA Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,
 A Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,
 A Ishiguro H.;
 RT "A novel brain-derived member of the epidermal growth factor family
 RT that interacts with ErbB3 and ErbB4.";
 RL J. Biochem. 122:675-680(1997).
 RN [2]
 RP SEQUENCE OF 109-868 FROM N.A. (ISOFORMS NRG2-ALPHA AND NRG2-BETA).
 RC TISSUE=Cerebellum;
 RX MEDLINE=97311397; PubMed=9168114;
 RA Chang H., Riese D.J. II, Gilbert W., Stern D.F., McMahon U.J.;
 RT "Ligands for ErbB-family receptors encoded by a neuregulin-like
 RT gene.";
 RL Nature 387:509-512(1997).
 CC -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
 CC RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
 CC RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
 CC ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE
 CC HETERODIMERIZATION WITH THE EGF RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
 CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
 CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS: NTAK-ALPHA1 (SHOWN
 CC HERE), NTAK-ALPHA2A, NTAK-ALPHA2B/NTAK-ALPHA2-1P, NTAK-BETA, NTAK-
 CC GAMMA, NRG2-ALPHA AND NRG2-BETA; ARE PRODUCED BY ALTERNATIVE
 CC SPLICING. THE ALPHA-TYPE AND BETA-TYPE DIFFER IN THE EGF-LIKE
 CC DOMAIN.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST PARTS OF THE BRAIN,

CC ESPECIALLY THE OLFACTORY BULB AND CEREBELLUM WHERE IT LOCALIZES IN
 CC GRANULE AND PURKINJE CELLS. IN THE HIPPOCAMPUS, FOUND IN THE
 CC GRANULE CELLS OF THE DENTATE GYRUS. IN THE BASAL FOREBRAIN, FOUND
 CC IN THE CHOLINERGIC CELLS. IN THE HINDBRAIN, WEAKLY DETECTABLE IN
 CC THE MOTOR TRIGEMINAL NUCLEUS. NOT DETECTED IN THE HYPOTHALAMUS.
 CC ALSO FOUND IN THE LIVER AND IN THE THYMUS. NOT DETECTED IN HEART,
 CC ADRENAL GLAND, OR TESTIS.
 CC -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, EXPRESSED IN THE BRAIN OF
 CC E11.5 EMBRYOS WHERE IT IS FOUND IN THE TELENCEPHALON, BUT NOT IN
 CC THE HINDBRAIN. NOT FOUND IN THE HEART. IN THE ADULT, FOUND IN
 CC BRAIN AND THYMUS.
 CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
 CC OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
 CC PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
 CC DIMERIZATION (BY SIMILARITY).
 CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
 CC DOMAIN (BY SIMILARITY).
 CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
 CC EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
 CC FORM (BY SIMILARITY).
 CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D89995; BAA23344.1; -
 CC EMBL; D89996; BAA23345.1; -
 CC EMBL; D89997; BAA23346.1; -
 CC EMBL; D89998; BAA23347.1; -
 CC EMBL; AB001576; BAA23348.1; -
 CC InterPro: IPR000561; -
 CC InterPro: IPR003006; -
 CC Pfam: PF00008; EGF; 1.
 CC Pfam: PF00047; IG; 1.
 CC PROSITE: PS00022; EGF_1; 1.
 CC PROSITE: PS01186; EGF_2; 1.
 CC Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 CC Transmembrane; Multigene family; Alternative splicing.
 CC PROPEP 1 127
 FT CHAIN 128 868
 FT CHAIN 128 428
 FT DOMAIN 128 429
 FT TRANSMEM 430 450
 FT DOMAIN 451 868
 FT DOMAIN 266 334
 FT DOMAIN 346 356
 FT DOMAIN 357 398
 FT DOMAIN 22 32
 FT DOMAIN 35 45
 FT DOMAIN 56 59
 FT DOMAIN 103 106
 FT DOMAIN 739 745
 FT DISULFD -273 327
 FT DISULFD 361 375
 FT DISULFD 369 386
 FT DISULFD 388 397
 FT CARBOHYD 33 33
 FT CARBOHYD 34 34
 FT CARBOHYD 163 163
 FT CARBOHYD 294 294
 FT CARBOHYD 362 362
 FT CARBOHYD 1 108
 FT VARSPLIC 220 222
 FT VARSPLIC 388 388
 C -> G (IN ISOFORM NTAK-GAMMA).

| | | | | |
|----|----------|---------|-----------|--|
| FT | VARSPLIC | 389 | 868 | MISSING (IN ISOFORM NTAK-GAMMA). |
| FT | VARSPLIC | 390 | 412 | NGFTGQRCLEKPLRLYMPDKQ -> VGYTGRCCQCFAMV |
| FT | VARSPLIC | 390 | 421 | NFS (IN ISOFORM NRG2-BETA). |
| FT | VARSPLIC | 414 | 421 | NGFTGQRCLEKPLRLYMPDKQKHLGFLKE -> VGYTGRCCQCFAMVFSK (IN ISOFORM NTAK-BETA). |
| FT | VARSPLIC | 414 | 439 | MISSING (IN ISOFORM NTAK-ALPHA2A AND ISOFORM NTAK-ALPHA2B). |
| FT | VARSPLIC | 440 | 868 | HLCPELKEELYOKRVLTITIGICVA -> SVLWDTPTGTGV |
| FT | VARSPLIC | 117 | 117 | SSOWSTSPSLDLN (IN ISOFORM NRG2-ALPHA). |
| FT | CONFLICT | 724 | 724 | MISSING (IN ISOFORM NRG2-ALPHA). |
| FT | CONFLICT | 724 | 724 | S -> F (IN REF. 2). |
| FT | CONFLICT | 724 | 724 | R -> H (IN REF. 2). |
| SQ | SEQUENCE | 868 AA; | 93776 MW; | 3C7D4D94DBE64DE2 CRC64; |

| | | | | |
|-----------------------|-------|-----------------|-------|--------------------------------------|
| Query Match | 3.4% | Score 109.5; | DB 1; | Length 868; |
| Best Local Similarity | 19.1% | Pred. No. 0.81; | | |
| Matches | 92; | Conservative | 57; | Mismatches 173; Indels 159; Gaps 20; |

| | | | |
|----|-----|---|-----|
| Qy | 31 | RPDSPRRTPKGKPHAQOP-----GRASAS-----DSSAPWSRSTDTGTLA----- | 71 |
| Db | 74 | RPAAPPEPRPQPPQPRSPAARRAAARAAAGMRRDPAPGSSMLLFGVSLACYSPSL | 133 |
| Qy | 72 | ---OKLAEEVPMVASYLY-----TGDShOLKRANCGRVELAGLPGKWPALASAHPSL | 122 |
| Db | 134 | KSVQDQAYKAPVVEGKVOGLAPAGGSSNSTREPPASGRVALVKVLDKWP----- | 184 |
| Qy | 123 | HRALDTLTHATNFLNMVLOSNKRSRONLO---DDLWYQALVMSLLREGEPSISRAAITF | 178 |
| Db | 185 | -----LRSGGLQREQIVSVGSCAPLERNQRYIFFL---EP--TEQPLVF | 223 |
| Qy | 179 | STDSLSAPAPQVFLQATREESRIILLODSSSAPHLANATLETWFHGLRRKWRPHLHRRG | 238 |
| Db | 224 | KT--AFAPVDPNGKNIKKEVGKILCTDCAT-----RPKLKKMK | 259 |
| Qy | 239 | PNOGPRGLGHSWRRKDGKSHFKWSPPYLECENGSKPGWLVTLSAI---YG--- | 291 |
| Db | 260 | SQTGEVGEKQSLKCEAAGNPQPSYRW-----FKDGKELNRSRDIRIKYNGRK | 308 |
| Qy | 292 | ---LQPNLVP-----EPRGMKVVDINLOKVIDQCSS--DGWFSGTHKCH----- | 331 |
| Db | 309 | NSRLQFNKVKVEDAGEYVCEAENILGKDTVRGLHVNVSVTTLSSWSGHARKCNETAKSY | 368 |
| Qy | 332 | -LNNSECHPIKGLGVLCAYECICKAGFYHGVLPVNNERRGPD---OHISGSTKDVSE | 387 |
| Db | 369 | CVNGGVYVYIEG-----INOLSKCPCNPFQRCLEKPLRLYMPDKQKHLGFLKE--AE | 423 |
| Qy | 388 | EAYVCLPCREGCPFCADDSPCFVQEDKYLRLAITSFQGLCMLDVFSMLVYVHPRKAKSI | 447 |
| Db | 424 | ELY-----QKRVLTITIGICVALLVGVVAYCKTKKQ | 457 |
| Qy | 448 | R 448 | |
| Db | 458 | R 458 | |

Search completed: November 1, 2001, 10:20:48
Job time: 381 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2001, 10:19:21 ; Search time 61.87 Seconds
(without alignments)
1280.923 Million cell updates/sec

Title: US-09-775-181-4
Perfect score: 3199
Sequence: 1 MGNMAYPLLLLLQLGLG.....YMTAVGMWSLVSDGLTIFQ 599

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_unclassified: *
13: sp_vertebrate: *
14: sp_virus: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 339 | 10.6 | 669 | 5 Q9VR40 | Q9vr40 drosophila |
| 2 | 238.5 | 7.5 | 176 | 5 Q9WIP4 | Q9wip4 drosophila |
| 3 | 161.5 | 5.0 | 976 | 5 Q9V485 | Q9v485 drosophila |
| 4 | 149 | 4.7 | 1677 | 5 Q9VKA3 | Q9vka3 drosophila |
| 5 | 132.5 | 4.1 | 868 | 13 Q73636 | Q73636 fugu rubrip |
| 6 | 130 | 4.1 | 1305 | 5 Q9VPS7 | Q9vps7 drosophila |
| 7 | 127.5 | 4.0 | 840 | 11 Q9Z0R8 | Q9z0r8 rattus norv |
| 8 | 125 | 3.9 | 977 | 13 Q9PWE1 | Q9pwe1 ictalurus p |
| 9 | 124.5 | 3.9 | 528 | 5 Q96954 | Q96954 geodia cydo |
| 10 | 120.5 | 3.8 | 879 | 11 Q9QVS2 | Q9qvs2 mus musculu |
| 11 | 120 | 3.8 | 738 | 5 Q9V4U3 | Q9v4u3 drosophila |
| 12 | 120 | 3.8 | 870 | 5 Q9H4T8 | Q9h4t8 caenorhabdi |
| 13 | 120 | 3.8 | 872 | 4 Q9H3N6 | Q9h3n6 homo sapien |
| 14 | 117 | 3.7 | 264 | 5 Q9WIP3 | Q9wip3 drosophila |
| 15 | 116 | 3.6 | 589 | 4 Q9UGS9 | Q9ugs9 homo sapien |
| 16 | 116 | 3.6 | 877 | 4 Q9UGT0 | Q9ugt0 homo sapien |
| 17 | 115 | 3.6 | 1199 | 11 Q9EPV6 | Q9epv6 mus musculu |
| 18 | 113.5 | 3.5 | 940 | 13 Q73635 | Q73635 fugu rubrip |
| 19 | 111.5 | 3.5 | 362 | 13 Q9PSY1 | Q9psyl carassius a |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 20 | 111.5 | 3.5 | 408 | 13 | Q93558 | O93558 carassius a |
| 21 | 110.5 | 3.5 | 3507 | 5 | Q23587 | Q23587 caenorhabdi |
| 22 | 110 | 3.4 | 540 | 5 | Q9YIA5 | Q9yla5 lymnaea sta |
| 23 | 107.5 | 3.4 | 877 | 13 | Q9PW88 | Q9pw88 carassius a |
| 24 | 107.5 | 3.4 | 2437 | 11 | O63725 | O63725 rattus norv |
| 25 | 107 | 3.3 | 383 | 11 | O70534 | O70534 rattus norv |
| 26 | 107 | 3.3 | 848 | 13 | Q93553 | Q93553 carassius a |
| 27 | 107 | 3.3 | 1956 | 4 | Q9Y5Y9 | Q9y5y9 homo sapien |
| 28 | 106 | 3.3 | 350 | 13 | Q93556 | Q93556 carassius a |
| 29 | 106 | 3.3 | 551 | 13 | Q9PWQ0 | Q9pwq0 fugu rubrip |
| 30 | 105 | 3.3 | 383 | 11 | O62779 | O62779 rattus norv |
| 31 | 104.5 | 3.3 | 3312 | 4 | Q9NY07 | Q9nyq7 homo sapien |
| 32 | 102.5 | 3.2 | 779 | 11 | O35269 | O35269 rattus norv |
| 33 | 102.5 | 3.2 | 797 | 4 | O95288 | O95288 homo sapien |
| 34 | 102.5 | 3.2 | 3680 | 5 | Q9VR08 | Q9vr08 drosophila |
| 35 | 101.5 | 3.2 | 864 | 13 | O73637 | O73637 fugu rubrip |
| 36 | 101.5 | 3.2 | 934 | 13 | Q9DER4 | Q9der4 gallus gall |
| 37 | 101 | 3.2 | 1152 | 11 | O70375 | O70375 mus musculu |
| 38 | 100.5 | 3.1 | 2160 | 3 | O13328 | O13328 magnaporthe |
| 39 | 100.5 | 3.1 | 2160 | 3 | O13488 | O13488 magnaporthe |
| 40 | 100.5 | 3.1 | 3313 | 11 | O88278 | O88278 rattus norv |
| 41 | 100 | 3.1 | 1062 | 11 | O60789 | O60789 mus musculu |
| 42 | 100 | 3.1 | 1993 | 5 | P90670 | P90670 aplysia cal |
| 43 | 99.5 | 3.1 | 250 | 13 | O73645 | O73645 fugu rubrip |
| 44 | 99.5 | 3.1 | 2471 | 4 | Q04721 | Q04721 homo sapien |
| 45 | 99.5 | 3.1 | 2471 | 4 | Q9H240 | Q9h240 homo sapien |

ALIGNMENTS

RESULT 1

| | | | | | |
|--------|---|--------|--------------|------|---------|
| Q9VR40 | ID | Q9VR40 | PRELIMINARY; | PRT; | 669 AA. |
| AC | Q9VR40; | | | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Created) | | | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last sequence update) | | | | |
| DE | 01-MAR-2001 (TREMBLrel. 16, Last annotation update) | | | | |
| DE | CG11923 PROTEIN. | | | | |
| GN | CG11923. | | | | |
| OS | Drosophila melanogaster (Fruit fly). | | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | | | |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | | | |
| OX | NCBI_TaxID=7227; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | SFRAIN-BERKELEY; | | | | |
| RX | MEDLINE=20196006; PubMed=10731132; | | | | |
| RA | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., | | | | |
| RA | Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., | | | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | | | |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | | | | |
| RA | Randall R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D., | | | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G., | | | | |
| RA | Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D., | | | | |
| RA | Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | | | |
| RA | Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | | | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P., | | | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., | | | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | | | |
| RA | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., | | | | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | | | |
| RA | Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., | | | | |
| RA | Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., | | | | |
| RA | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | | | |
| RA | Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houch J., | | | | |
| RA | Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., | | | | |
| RA | Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | | | | |
| RA | Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., | | | | |
| RA | Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., | | | | |
| RA | Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., | | | | |
| RA | Merkulov G., Milshina N.V., Mobarry J., Morris J., Moshrefi A., | | | | |

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE GLU-RA PROTEIN.
GN GLU-RA OR CG11144.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
EMBL: AE003846; AAF59402.1; -;
FlyBase: FBgn0019985; Glu-RA.
DR InterPro: IPR000337;
DR InterPro: IPR001828;
DR Pfam: PF00003; 7tm3; 1.
DR PRINTS: PR001094; ANF_receptor; 1.
DR PROSITE: PS00979; G_PROTEIN_RECF_F3_1; 1.
DR PROSITE: PS00980; G_PROTEIN_RECF_F3_2; 1.
DR PROSITE: PS00981; G_PROTEIN_RECF_F3_3; 1.
SO SEQUENCE 976 AA; 108485 MW; 43A0E1F918EDACC4 CRC64;

Query Match 5.08; Score 161.5; DB 5; Length 976;
Best Local Similarity 20.18; Pred. No. 6.2e-05;
Matches 101; Conservative 80; Mismatches 179; Indels 143; Gaps 21;

Qy 145 SREONLQDDLDWYQALV-----SLLEGPSISRAAITSTDS-LSAPAPQVLOATRES 199
Db 295 AKRNLSPFWHIASDGGWKQKLEGLEDIAEGNITVELSEIADFDRIYMQLTPTN 354

Qy 200 R-----ILLQDLSSSAPHLANATLETETWHLGRKWRPHLHRRGNQGPGRG 245
Db 355 QRNPFABEYWEDETFNCVLTSLSVK-PDTSNSANSTDNKIGVKAKTE----- 399
Qy 246 LCHSWRRKDGGLGGD---KSHFKWSPPY-----LECENG 275
Db 400 CDDSVRLSEKVGVEQESKTQFVVDVAYAFAYALHNLHNDRCNTQSDQTEFRKHLQSESV 459
Qy 276 SYKPGWLVLTLSSA-----IYGLQPNLVPEFRGVMKVDI-NL 310
Db 460 WYRKISTDTKSOACPDMDANYDKGEFYNVLLNVSIDLAGSEVKFDROGDGLARYDILNY 519
Qy 311 QKVIDQSSDG-----WFSGTHKCHLNNSECMPTKGLGVLGAVECTCKAGFYHPG 362
Db 520 QR-----QENSQGYQYKVIKWFNG---LQLANSETVVMKETEQTPTSACSLPCEVGM--- 568
Qy 363 VLPVNNRRRRPDQHI-----SGSTKDVSEEAIVCLPCREGCPFCADDSPCFVQEDKYL- 417
Db 569 -----IKKQGDYCCWICDSCSEFEYVYDEFTCKDQGGPLWYADKUSCFALDIQYMKW 622
Qy 418 -----LATISFQGLCMLLDFVSMVYVYHFRK---AKSTRASGLIILETILFGSLLY 466
Db 623 NSLPALIPMAIAIF-GIAL-----TSIVILVFAKNHDTPLVRASGRELSYTLFGILVCY 676
Qy 467 FPVVILYEPSTFRGILLRWLRLLGFAVYGVTVLKLHRLVKVF--LSRTQRIPTYMTGG 524
Db 577 CNTFALIAKPTIGSCVLQRFQGVGFSIYSALLTKTNRISRFHSKASKAQRLLKYSIQ 736
Qy 525 R-----VMEMLAV-ILLVVFWEFLI 542
Db 737 SQVVTTSLLIAIQVLITWIMWV 759
RESULT 4
Q9VKA3 PRELIMINARY; PRT: 1677 AA.
AC Q9VKA3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE CG17215 PROTEIN.
GN CG17215
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fogle C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
EMBL: AE003846; AAF59402.1; -;
FlyBase: FBgn0019985; Glu-RA.
DR InterPro: IPR000337;
DR InterPro: IPR001828;
DR Pfam: PF00003; 7tm3; 1.
DR PRINTS: PR001094; ANF_receptor; 1.
DR PROSITE: PS00979; G_PROTEIN_RECF_F3_1; 1.
DR PROSITE: PS00980; G_PROTEIN_RECF_F3_2; 1.
DR PROSITE: PS00981; G_PROTEIN_RECF_F3_3; 1.
SO SEQUENCE 976 AA; 108485 MW; 43A0E1F918EDACC4 CRC64;


```
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG3022 PROTEIN.
GN CG3022.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003588; AAF51465.2;
DR FlyBase: FBgn0031275; CG3022.
DR InterPro: IPR000005;
DR InterPro: IPR000337;
DR InterPro: IPR001828;
DR Pfam: PF000003; 7tm.3; 1.
DR Pfam: PF01094; ANF_receptor; 2.
DR PRINTS: PR00248; GPCRMR.
DR PROSITE: PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
DR SEQUENCE 1305 AA; 143588 MW; A9C3C85307650450 CRC64;

Query Match 4.1%; Score 130; DB 5; Length 1305;
Best Local Similarity 19.2%; Pred. No. 0.04;
Matches 87; Conservative 68; Mismatches 164; Indels 134; Gaps 20;

QY 181 DSLAP-APOVFLOATREESRILLQDLSSSAPH-----LANATLETWFHGLRRKWRP 232
DB 383 ESMGAPWPDQRTACSNHELOLAVENLTVSTHNSVGNVSYGLNNHMFNSOLRQSA 442
QY 233 HLHRRGNQGRGLGHSW-----RRKDLGGDKSHFKWSPYPLECGNSY 277
DB 443 QFH-----GODGFGSGYGRISIAATQSDSRRRRRRGVGGTSGGHLFPPEAISQAPQY 496
```

```
QY 278 KPGWLVTLSAIYGLQPNLVPFRGVKMKVDINLOKVIDIDQSSDGFSGTHKCHLNHSEC 337
DB 497 DAVWAIALAL-----RAAEHWH-----RRNEEQ 519
QY 338 MPIKGLGCVLGG--AYECICKAGFYH-PCVLPVNNFRRRGGPDQHISGSP-----KDYSEE 388
DB 520 SKLDGFDYTRSDMAWEFLQMGKHLFLGVSGPVSF--SGPDR--VGTTAFYQIORGLLEP 575
QY 389 AYVCLPCREG-----CPFC-----ADDSFCFVQEDKYLRLAIIISF 423
DB 576 VALIYPATDLDALDFRCPRCPVKNHSGOVPIAKRVKLRVATAPL-----AFYTIATLSS 630
QY 424 QGLCMLDLFVSMVYVYHFRKAKSTRASGLILLETILFGLSLLYPVPVILYEPST-----478
DB 631 VGIALAIAFLAFNL--HFRKLKAIKLSPLKSNITAVGCIYVATVILLGLDHTLPSAE 688
QY 479 --FRCLLRWALL--GFATVYGVTVTLKLRVLAFLVLSRTAQ-----RIPYMTGG 524
DB 689 DSFATVCTARVYLLSAGFSLAGSMFAKTYRVHRIF--TRTGSVFKDKMLQDILILVGG 747
QY 525 RVMRLAVILLVWFVFLIGTSSVCQNLEKQIS 557
DB 748 ---LLIVDALLVTLVWVTPDPMERHLNLTLEIS 777

RESULT 7
QY20R8 PRELIMINARY; PRT; 840 AA.
AC QY20R8;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE TASTE RECEPTOR TR1 (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=99159821; PubMed=10052456;
RA Hoon M.A., Adler E., Lindemeier J., Battey J.F., Ryba N.J.P.,
RA Zuker C.S.;
RT "Putative mammalian taste receptors: a class of taste-specific GPCRs
RT with distinct topographic selectivity."
RL Cell 96:541-551(1999)
DR EMBL: AF127389; AAD18069.1;
DR InterPro: IPR000337;
DR InterPro: IPR001828;
DR Pfam: PF000003; 7tm.3; 1.
DR Pfam: PF01094; ANF_receptor; 1.
DR PRINTS: PR00248; GPCRMR.
KW Receptor.
FT NON_TER 840 840
SQ SEQUENCE 840 AA; 93496 MW; 1FCFB7EFC6B45DB CRC64;

Query Match 4.0%; Score 127.5; DB 11; Length 840;
Best Local Similarity 22.2%; Pred. No. 0.037;
Matches 60; Conservative 48; Mismatches 89; Indels 73; Gaps 14;

QY 306 VDINLQ-----VDIDQSSDGFSGTHKCHLNHSECMPKIGLFLVGAYECI-CK 355
DB 475 LDINKTKIOWHGKNNQVPVSVCTTD-CLAGHHRVVVGSHHC-----CPECVCE 522
QY 356 AGFTHPGVLPVNNFRRRGGPDQHISGSKVDSEAYVCLPCL--REGCPFCADDSPCFVQED 413
DB 523 AGTF-----LN-----MSLHICQPCGTEEWAP--KESTCFPRTV 556
QY 414 KYL-----RLAIIISFOGLCMLDGF-VSMVYVYHFRKAKSTRASG-----LILLETILFGS 462
DB 557 EFLAWHEPISLVLIANAANTLLLLLVGTAGLPAWHFH-TPVVRSAAGRLCFLMLGLSVAGS 615
```

QY 463 LLLYPVWLYPEPSTFRICLLRWARKLGFATVGVVTLKLHVLKVLFSRTAQRIPYMT 522
DB 616 CSFY-----SFFGEPVTPACLLRPLSLGFAIFLSCLTIRSFOLVIFRSTKVPFTYRT 671
QY 523 -----GGRVRLMVAIVLVWFNFWLWMT 545
DB 672 WAQNHAGLGFVIVSVTHILLICLTWLVMT 701

RESULT 8
Q9PWE1 PRELIMINARY; PRT; 977 AA.
AC Q9PWE1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 16, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR (FRAGMENT).
GLUR.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Siluriformes; Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Hellmich H.L., Micci A.M., Sanchez X., Christensen B.N.;
RT "Molecular cloning, functional expression and localization of a novel
RT metabotropic glutamate receptor linked to calcium mobilization from
RT the catfish retina."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF076473; AAD47893.1;
DR InterPro; IPR00337;
DR InterPro; IPR001828;
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
KW Receptor.
FT NON_TER 977 977
SQ SEQUENCE 977 AA; 108759 MW; BD06A67E240751CE CRC64;

Query Match 3.9%; Score 125; DB 13; Length 977;
Best Local Similarity 25.4%; Pred. No. 0.072;
Matches 45; Conservative 30; Mismatches 74; Indels 28; Gaps 7;
QY 392 CLPCR-----EGCPF-----CADDSPC-----FVQEDKYLRLAII-SFGGLCMLLD 431
DB 572 CEPDGYQYQVGLTCEMCPFDMRPTANHTACTPTPIIKLDWHSWAVVPMFLAILGIAA 631
QY 432 FYSMLVWY-HFKAKSIRASGLILLETILFGLSLLYFPVVLVYFEPSTERCILLRWALL 490
DB 632 TLISVIVFRNDTPIVRASGRELSTVLTLGLIYLTITFIAGPNTVVCALRLLGL 691
QY 491 GFATVGVVTLKLHRLVKVF-----LSRTAQRIPYMTGGRVRLAV---ILLVWF 540
DB 692 GNCITYSAMLTKNRIYRIFEGKKSVTAPKFIPTISQLVITFILVSFOVIGVIFW 748

RESULT 9
O96954 PRELIMINARY; PRT; 528 AA.
AC O96954;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2001 (TREMBLrel. 16, Last annotation update)
DE METABOTROPIC GLUTAMATE GABA-LIKE RECEPTOR.
GN MGR1.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;

OC Astrophorida; Geodiidae; Geodia.
OX NCBI_TaxID=6047;
RN [1]
RP SEQUENCE FROM N.A.
RA Perovic S., Prokic I., Krasko A., Mueller I.M., Mueller W.E.G.;
RT "Origin of neuronal receptors in Metazoa: cloning of a metabotropic
RT glutamate/-like receptor from the marine sponge Geodia cydonium."
RL Cell Tissue Res. 0:0-0(0).
DR EMBL; Y17211; CAA76688.1;
DR InterPro; IPR00337;
DR Pfam; PF00003; 7tm_3; 1.
KW Receptor.
SQ SEQUENCE 528 AA; 58641 MW; OB99D8357FEAB3B1 CRC64;

Query Match 3.9%; Score 124.5; DB 5; Length 528;
Best Local Similarity 32.6%; Pred. No. 0.036;
Matches 42; Conservative 20; Mismatches 52; Indels 15; Gaps 5;
QY 415 YLRALISFQGLCMLDFVSMVYHFRKAKSIRASGLILLETILFGLSLLYFPVVLVYF 474
DB 141 YVALAV---GGLVFAIVCVFETVI--FRKKLIRLSSPNLVILGILGAILYFNVTLVI 195
QY 475 EPST-----FRCILLRWALLGFATVGVVTLKLHRLVKVLSRTAQRIPYMTGGRV 528
DB 196 -PTDTVIAAILCINPMLTSLGSLCYCTILAKTIRWFI--NKRVPVSVTKSIVIK 251
QY 529 MLAVILLVV 537
DB 252 DYALALFVV 260

RESULT 10
Q9QYS2 PRELIMINARY; PRT; 879 AA.
ID Q9QYS2;
AC Q9QYS2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2001 (TREMBLrel. 16, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 3 PROTEIN.
GN GRM3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV.
RX MEDLINE=20012997; PubMed=10544282;
RA Minoshima T., Nakanishi S.;
RT "Structural organization of the mouse metabotropic glutamate receptor
RT subtype 3 and its regulation by growth factors in cultured cortical
RT astrocytes".
RL J. Biochem. 126:889-896(1999).
DR EMBL; AF170701; AAF06741.1;
DR EMBL; AF170697; AAF06741.1; JOINED.
DR EMBL; AF170698; AAF06741.1; JOINED.
DR EMBL; AF170699; AAF06741.1; JOINED.
DR EMBL; AF170700; AAF06741.1; JOINED.
DR InterPro; IPR00337;
DR InterPro; IPR001828;
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
KW Receptor.
SQ SEQUENCE 879 AA; 99113 MW; F3A8B26CE96679EF CRC64;

Query Match 3.8%; Score 120.5; DB 11; Length 879;
Best Local Similarity 18.2%; Pred. No. 0.15;

```
Matches 91; Conservative 65; Mismatches 185; Indels 159; Gaps 21;
QY 138 VMLGNSKRE-----QNIQDLDWQYALV-----SLLEGESISRRAITFTDLSAPAP 188
Db 274 LFMRSDDSRILIAAASRVNASFTVWASDGMGAQESIVKGSBHVAIGATTLELAS----- 327
QY 189 QVFLQATREESRILLQDLSAPHLANATLETETFWHGLRRKWRPHILHRRGNQPRGLGH 248
Db 328 ----HPVRQFDR-YFQSLNPNYNNH-----RNPWFRDP----- 354
QY 249 SWRRK--DGLGCDKSHKSPPYLECEGSKYP-----GWLVTLSAAYGLQPNLVP 298
Db 355 -WEQKFCQSLQNRHRCIDKHLAIDSSNYEQESKIMFVNVNAYAMAHLRKQRTLCP 413
QY 299 EFRGMKVDINLQKVDIQDSSDQWGFSTHCKHLNNSECMPKIGLGFVLGAYECICKAGF 358
Db 414 ---NTTKCDAMKILDKGKLYLLKINFTAPEN-----PNKGA-----DSTVKFDT 458
QY 359 YHPGVLPNNRRRGPDQIHSGS-----TKDVSEB-- 388
Db 459 YGDGMGRYNFEN-----FOHIGGKYSYLVKGHWAETLYLDVDSIHWSRNVPTSCSDPCA 514
QY 389 -----AYVCLPCRE-----GCPRFC-----ADDSPCFVQEDKYLK--- 417
Db 515 PNEMKNQPGDVCCWICIPCEPYEYLYDEFTCMCGGQWPTADLSGCYNLPEDYIRWED 574
QY 418 -----LATISFOGLCMLDFVSMVLVYHFRKAKSIRASGLILLETILFGSLLLPVP 469
Db 575 AWATGPVTIACLGPMCTCIVI-----TVFIKHNTPLVKASGRELCYILLFGVSLSYCMT 629
QY 470 VILFEFSTFCILLRWARLLGFTVGTVTLLKRLVUKVP--LSRTRAQRIPYMT--GGRV 526
Db 630 FFFAKSPVICARRRLGLGFSFATCYSALITKTNCIARIFDGVKNGAQRKPFISPSQV 689
QY 527 MRMLAVIL---LVVFWFLI 542
Db 690 FICGLILQIVMVSVWDLIL 709

RESULT 11
Q9V4U3 PRELIMINARY; PRT: 738 AA.
ID Q9V4U3
AC Q9V4U3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG8692 PROTEIN.
GN CG8692.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtova K.C., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

```
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel M.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR FlyBase: FBgn0033298; AAF59082.1; -.
DR ENBL: AE003837; AAF59082.1; -.
DR InterPro: IPR000337; -.
DR InterPro: IPR001828; -.
DR Pfam: PF00003; 7tm_3; 1.
DR Pfam: PF01094; ANF_receptor; 1.
DR PRINTS: PR00248; GPCRMR.
DR PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
SQ SEQUENCE 738 AA; 83179 MW; 0E6620337B059EAC CRC64;
```

Query Match 3.8%; Score 120; DB 5; Length 738;

Best Local Similarity 18.8%; Pred. No. 0.13;

Matches 106; Conservative 63; Mismatches 198; Indels 196; Gaps 26;

```
QY 70 LAQKLAEEVPMDVASYLYTGDS-----HQLKRANCSGRYELAGLPKWPALASAPSLH 123
Db 62 IVQKLLTK-PRAGAIIFGSDQEVQRQVMRAVRANATGSFWSIGSDG-WSA----- 110
QY 124 RALDTLTHATNFLNVLQSNKSRQNLQDLDWQYALVSLLEGESISRRAITFTDLSL 183
Db 111 -----RNLVSD-----DYEPEVG-----TLVSVQPO 131
QY 184 SAPA---POVFLOATRESRILLQDLSAPHLANATLETETFWHGLRRKWRPHILHRRGN 240
Db 132 ANPYRGPEEYFLSLTVENNQ-----RNPWFVEF---WEDHFCQRYPG 170
QY 241 QGPRGLGHSWRKDKGLGDKSHFKWSPY-----LECNGSYKPGWLVTLSA 288
Db 171 -----STSTPYNNYTKQCTTKERLSRONTDFE-DQLQFVSDA 206
QY 289 I-----YGLQ-----PNLVPFRGVMKVDI--NLQKVDIDQCSSDGW-FSGT----- 327
Db 207 VMAFAYALRDMHRDLCGGSPCLCEAMKPTKGADLLKYLRKVEFGLSGDEFEDGNGDGP 266
QY 328 ---HKCHLNNSECMPKIGLGFVLGAYECICKAGFYHPGVLPWN-----NFRRRGPDQHS- 379
Db 267 ARYNIHFQKQSA-----GOYHWV-KVGETEGELRLNMTVEFKRLSPKPPESV 315
QY 380 -----GSTKDVSEAYVCLPC-----REGCPFC-----ADDSPCFVQE 412
Db 316 CSLPCLVGAQAKYVEGESCCWHCFNCTTYQIRHPDDETHCKLCKLGLTLPDAHKKQYCRIP 375
QY 413 DKYLR-----LATISFOGLCMLDFVSMVLVYHFRKAKSIRASGLILLETILFGSLLLY 466
Db 376 EYLRPESAWAIGAMAFSATGILVTLFVGMVFVRHNDTPIVRASRELSYILLAGIFCY 435
QY 467 FPMVILYEPSTFCILLRWARLLGFTVGTVTLLKRLVUKVLS--RTAQRIPYMTGG 524
Db 436 GVTFLVLUKPTNIVCAIORFGVGCTFVVYAAALLTKTNRIARIFRAGKQSAKRPFSIPK 495
```

```
QY 525 RVMRLAV-----ILLVWFLLI 542
DE : : |
GN : : | : |
Db 496 SOLVICACLSVQVLINGVMVJ 518

RESULT 12
Q9N4T8 PRELIMINARY; PRT: 870 AA.
ID Q9N4T8
AC Q9N4T8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Y4C6A.2 PROTEIN.
GN Y4C6A.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Layman D., Graves T., Yoakum M.;
RT "The sequence of C. elegans cosmid Y4C6A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006791; AAF60738.1; -.
DR InterPro; IPR000337; -.
DR InterPro; IPR001828; -.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 2.
DR PRINTS; PR00248; GPCRMR.
SQ SEQUENCE 870 AA; 97727 MW; 2FCE2F871D6B1EF9 CRC64;

Query Match 3.8%; Score 120; DB 5; Length 870;
Best Local Similarity 22.7%; Pred. No. 0.16;
Matches 49; Conservative 35; Mismatches 102; Indels 30; Gaps 5;

QY 390 YVCLPCR-----EGCP-----FCADDSPCFQEDKYLRLAIIISFQGLCM 428
Db 566 WACIPCDTSTSIHNETSCCEACVGMVDPRTLHFCVPIPPVSMQWDITWSLIPAFTSLGI 625

QY 429 LIDFVSMVYVIFRAKSTRASGLILLETILFGSLLYFPVILYFEPSTFCILLRAR 488
Db 626 ASTIFVSVFLKSNTPVIMASGRELICYMMSGIGMXYLTFLVLSQPTVITCSMTFILM 685

QY 489 LIGFATVYGTVTLKLRHLVLFSLRTAQRIPYMTG---GRVMRLAVILLVVF-WFLIG 543
Db 686 GLSMNIAIYAAITKTNRLARVFPDSAQRPRTFKAQVIGCMGVSLIGTFFVILFD 745

QY 544 WTSVVCQNLEKQISLIGQKTSDH-----LIFNMCLI 575
Db 746 PPGTMIVFPTRTEAVLTCKATTSHLLISLLYNILLI 781

RESULT 13
Q9H3N6 PRELIMINARY; PRT: 872 AA.
ID Q9H3N6
AC Q9H3N6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
```

```
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR TYPE 2.
GN HMG2LUR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yasuyuki F., Akiko J.;
RT "Structure and polymorphisms of the human metabotropic glutamate
RT receptor type 2 (hmg2lur2) gene: Analysis of association with
RT schizophrenia.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045011; BAB19817.1; -.
KW Receptor.
SQ SEQUENCE 872 AA; 95567 MW; 801976D034AA8100 CRC64;

Query Match 3.8%; Score 120; DB 4; Length 872;
Best Local Similarity 26.2%; Pred. No. 0.16;
Matches 49; Conservative 28; Mismatches 70; Indels 40; Gaps 9;

QY 390 YVCLPCR-----CADD-----SPCFQEDKYLRLAIIIS 422
Db 520 WLCIPQC---PYEYRLDEFTCADCGLYWPNASLTGCFELPQEIYRWDAWAVGPVTIAC 576

QY 423 FQGLCMILDFVSMVYVIFRAKSTRASGLILLETILFGSLLYFPVILYFEPSTFCRI 482
Db 577 LGALATL--FVLGVFVRH-NATPVVRASGRELICYLLGGVFLCYCMTFFIAKPSTAVCT 633

QY 483 LLRWARLLGFATVYGTVTLKLRHLVLF--LSRTAQRIPYMT--GGRVMRLAVI----LL 535
Db 634 LRRLLGTAFSVCYCSALLTKTNRIARIFGGAREGAQRPRFISPASQVAICLALISQLLI 693

QY 536 VVFWFLI 542
Db 694 VVAVLVV 700

RESULT 14
Q9WLP3 PRELIMINARY; PRT: 264 AA.
ID Q9WLP3
AC Q9WLP3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CG18678 PROTEIN.
GN CG18678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
FT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RA EMBL: AE003461; AAP47013.1; -.
RA FlyBase: FBgn0040664; CG18678.
RA InterPro: IPR000337; -.
DR Pfam: PF00003; 7tm_3; 1.
DR PRINTS: PR00248; GPCRMR.
SQ SEQUENCE 264 AA; 29532 MW; CAC5623C52703642 CRC64;

Query Match 3.7%; Score 117; DB 5; Length 264;
Best Local Similarity 26.1%; Pred. No. 0.064;
Matches 29; Conservative 23; Mismatches 47; Indels 12; Gaps 3;

QY 472 LYEPSTFRCLLWARLLGFATYGVGTTLKHLRVLKVFLSRTAQRIPIYMTGGVRMRLA 531
Db 1 MIPNPLTCTARLWLREIGFSLYTGALMLKTRISVIFVRSAKAV-ITDAALLRLG 59
QY 532 VILLVFWFLIGWTSVQCNLEKQISL--ICQKTSDLHLENMCLIDRWYD 580
Db 60 IICGAI-----GTCLLVRLVSPDPVVVGRVTTADDLKAFLCKTDWDY 101

RESULT 15
Q9UGS9
ID Q9UGS9 PRELIMINARY; PRT; 589 AA.
AC Q9UGS9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DU 69B13.1 (GLUTAMATE RECEPTOR, METABOTROPIC 1 (BETA ISOFORM))
(FRAGMENT).
GN GRM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035698; CAB65992.1; -.
DR InterPro: IPR000337; -.
DR InterPro: IPR001828; -.
DR Pfam: PF00003; 7tm_3; 1.
DR Pfam: PF01094; ANF_receptor; 1.
DR PRINTS: PR00248; GPCRMR.
DR PROSITE: PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE: PS00981; G_PROTEIN_RECEP_F3_3; 1.
KW Receptor.
FT NON_TER 1 1
SQ SEQUENCE 589 AA; 66304 MW; BEE9089D87293243 CRC64;

Query Match 3.6%; Score 116; DB 4; Length 589;
Best Local Similarity 22.1%; Pred. No. 0.22;
Matches 49; Conservative 34; Mismatches 95; Indels 44; Gaps 6;

QY 357 GFYHPGVLPVNNFRRGPDQHIISGSTDVDSEE-----AYVCLPCREG- 398
Db 181 GTWHEGVNLNDDYKIQ---MNKSGVRSVCSECLKGQIKVIRKEVSCCWCICTACKENE 237
QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMLV 438
Db 238 YVQDEFTCKACDLGWNPNADLTGCEPIPVRYLEWSNIESIIAIFSCGLIYLVTLFVLIF 297
QY 439 YHFRKAKSIRASGLILLETILFGLSLLYFPVVIYLFPPSTFRCLLRWARLLGFATYGT 498
Db 298 VLYRDTPVWRKSSSRELCYIILAGIFLGVVCPFTLIAKPTTTSCTCYLQRLVGLSSAMCYSA 357
QY 499 VTLKHLRVLKVFL---SRTAQRIPYMTGGVRMRLAVILLVW 537
Db 358 LVTKNRIARILAGSKKKIKTRKPRFMSAWAQVIAIISLISV 399

Search completed: November 1, 2001, 10:19:24
Job time: 352 sec

200.


```
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Ballault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 877)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 877)
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES             Location/Qualifiers
     source            1..877
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="025L08"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG025Df04SPI-end ;
                     PUC-ori"

BASE COUNT  177 a  277 c  267 g  151 t    5 others

alignment_scores:
    Quality: 293.00      Length: 87
    Ratio: 4.127        Gaps: 0
    Percent Similarity: 81.609   Percent Identity: 65.517

alignment_block:
US-09-775-181-4 x CNS03H19/rev ..

Align seg 1/1 to reverse of: CNS03H19 from: 1 to: 877

504 HisArgValLeuLysValPheLeuSerArgThrAlaGlnArgIleProty 520
577 CACAGGCTCTGAAGCTCTTCTTCCCGCACGCGCCAGAGATCCCTA 528
520 rMetThrGlyGlyArgValMetArgMetLeuAlaValIleLeuValV 537
527 CATGACAGCTGGCGCTCTGCGCTGTGCGCTGCTATCCTCGTGGTGG 478
537 alPheTrpPheLeuIleGlyTrpThrSerSerValCysGlnAsnLeuGlu 553
477 TCYGTCTGTTCTCTGGCGCTGTGACGTGCGCTGCTGCGAGACCCCGAC 428
554 LysGlnIleSerLeuIleGlyGlnGlyLeuThrSerAspHisLeuIlePh 570
427 AGGAAGCGCGCTCATCGACGTGGGTCCACGCCCGCGGCTGCAGTT 378
570 eAsnMetCysLeuIleAspArgTrpAspTyrMetThrAlaValGlyMetT 587
377 CGCCACCTGCTGCTGACCGCTGGGACTACATGATGGCTGTGGGTAAGC 328
587 rpSerLeuVal 590
327 ACCGCTCTCTG 317

seq_name: gb_gss34:A2903235

seq_documentation_block:
LOCUS AZ903235 483 bp DNA GSS 05-MAR-2001
DEFINITION RPCI-24-187D3.TJ RPCI-24 Mus musculus genomic clone RPCI-24-187D3,
DNA sequence.
```

```
ACCESSION AZ903235
VERSION AZ903235.1 GI:13222180
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 483)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.W.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-187D3.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.chofi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 187 row: D column: 3
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..483
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RPCI-24-187D3"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site.1: BamHI; Site.2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT  98 a  138 c  152 g  95 t

alignment_scores:
    Quality: 282.00      Length: 188
    Ratio: 2.474        Gaps: 4
    Percent Similarity: 60.638   Percent Identity: 34.574

alignment_block:
US-09-775-181-4 x AZ903235 ..

Align seg 1/1 to: AZ903235 from: 1 to: 483

1 MetGlyAlaMetAlaTrpProLeuLeuLeuCysLeuLeuAlaGlnLe 17
32 ATTTCCTCTCTTGGGGCTCCAGTAGTTGCTGTCTCTCTGTAGT.. 79
17 uGlyLeuGlyAlaValGlyAlaSerArgAspProGlnGlyArgProAsp 34
80 .....GGGCTCTAGGGAGCCAAAGGCCCTTGCCTCTCTGCCCTCCAC 122
34 erProArgGluArgThrProLysGlyLysProHisAlaGlnProGly 50
123 TACCGTCCCAAGCCAAAGCATCTGAGCCCATGTGATGCCCCCAAG 172
51 ArgAlaSerAlaSerAspSerSerAlaProTrpSerArgSerThrAspG 67
173 GGGGCTGAAGCAGCC..... 187
```

```
67 yThrIleLeuAlaGlnLysLeuAlaGluValProMetaspValAlas 84
188 .....CTGG 191
84 erTyLeuTyThrGlyAspSerHisGlnLeuLysArgAlaAsnCysSer 100
.....CTGG 100
192 CCTTCTCTACTCTGAGAGTGTCAACGGCTGTCAGGAGCAAACTGCAGT 241
101 GlyArgTyrgluLeuAlaGlyLeuProGlyLysTrpProAlaLeuAlase 117
.....CTGG 117
242 GAGAAATATGAAGTCGTGGGGCGGAAGCAAA.....GCCGG 279
117 rAlaHisProSerLeuHisArgAlaLeuAspThrLeuThrHisAlaThrA 134
.....CTGG 134
280 GTGCCCCCAGCTTCACAGAGACGCGGGCACCTGGCCCCAGGCTGCCA 329
134 snPheLeuAsnValMetLeuGlnSerAsnLysSerArgGluGlnAsnLeu 150
.....CTGG 150
330 ACTTCTCAACATGCTCTACAGCCACAGCATCCGGGAGTCGAGCGTG 379
151 GlnAspLeuAspTrpTyrglnAlaLeuValTrpSerLeuLeuGluG1 167
.....CTGG 167
380 GAGGAGGAGCGTGGAGTGTATCAGCGCTGTTCGACGTGTGGCAGAGGG 429
167 yGluProSerIleSerArgAlaAlaIleThrPheSerThrAspSerLeu 184
.....CTGG 184
430 AGACCCGAAGCCTACAGGGCTCTGCTGACCTTAAC..... 466
184 erAlaProAlaPro 188
.....CTGG 188
467 .....CTGGCACC 475
```

seq_name: gb_gss4 :CNS0304C

```
seq_documentation_block: 978 bp DNA GSS 17-MAY-2000
LOCUS CNS0304C Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 041G06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL252885
VERSION AL252885.1 GI:7973897
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 978)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billaut,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 978)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 978)
Genoscope.
Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL This sequence is a single read and was generated as part of a large
REFERENCE scale clone-end sequencing project of the tetraodon nigroviridis
AUTHORS genome. For more information, please take a look at
JOURNAL http://www.genoscope.cns.fr/Tetraodon.
COMMENT
FEATURES
source Location/Qualifiers
1..978 /organism="Tetraodon nigroviridis"
```

```
/db_xref="taxon:99883"
/clone="041G06"
/note="Genoscope sequence ID : COBG041BD03LP1-end : T7"
BASE COUNT 176 a 267 c 344 g 188 t 3 others
ORIGIN
alignment_scores:
Quality: 235.50 Length: 286
Ratio: 1.869 Gaps: 14
Percent Similarity: 44.056 Percent Identity: 26.923
alignment_block:
US-09-775-181-4 x CNS0304C/rev ..
```

```
Align seg 1/1 to reverse of: CNS0304C from: 1 to: 978
182 SerLeuSerAlaProAlaProGlnValPheLeuGlnAlaThrArgGluG1 198
|||||
956 TCCTCCCGCACCCCGCGGCAAGGTG..... 930
198 uSerArgIleLeuLeuGlnAspLeuSerSer.....AlaProHisL 213
|||
929 .CAGACCATCGTCTCCAGGACCTGTCCAAGGCTGGGACACGCTGCATC 881
213 euAlaAsnAlaThrLeuGluThrGluTrpPheHisGlyLeuArgArgLys 229
|||
880 TGCCCGCTCCCGCGGACGACAGCTGGTTCAGGAGT.....TTWAAA 837
230 TrpArgProHisLeuHisArgArgGlyProAsnGlnGlyProArgGlyLe 246
|||
836 TTCCCGCGG.....CCCAATCAG...CCGACAGCCTT 808
246 uGlyHis.....SerT 250
|||
807 GTCCAAACGGGTGCTCTTAAGACCTCAGCACCTGGACACGCCAAGT 758
250 tpArgArgLysAspGlyLeuGlyGlyAspLysSerHisPheLysTrp... 265
|||
757 GGGGCGGGGGGACAGCTACGTGACCAACCGCGGCGGCGCTGGGCC 708
266 Ser.ProProTyrrLeuGluCysGluAsnGlySerTyrrLysProGlyTrpL 282
|||
707 AGCGCCCCCGCTCTCGACTGCCAGACGGCGCTTCGTGCTGCTGCTGGA 658
282 euValThrLeuSerSerAlaIleTyrrGlyLeuGlnProAsnLeuValPro 298
|||
657 CGCTCACCTGTTCATGCCCTTCTAGGGCTCAAGCCCGACCTGACACCG 608
299 GluPhe..... 300
|||||
607 GAGTTCAGTTCAGGTCTCTCCACACACACACACACACACACACATC 558
300 ..... 300
557 ACACACAGGAAGCACCCCTTCCAAATGATTGAATTTGAATAATGAGGCC 508
300 ..... 300
507 TGACGGAGGCTCAGCGGGCTGTCCCCCTCCNCTCCGGTCCGTATGT 458
301 .....ArgGlyValMetLysValAspIleAsnLeuGlnLysValas 314
|||||
457 CTGTGCGTTAGAGGCTGATCGCGCTGGAGCTCAACATCCAGACATCGA 408
314 pIleAspGlnCys....SerSerAspGlyTrpPheSerGlyThrHisLysC 330
|||||
407 CTGTGACCATGTCAGCGGGGGGAGCGGTGTTCCGCCACACCCAGAGT 358
330 yHisLeuAsnAsnSerGluCysMetProIleLys..... 341
|||
357 GC.....AACCGCACCATGAGGTAAAGAAAGCCCTGACCCCGAG 317
```


Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1096)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1096)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1096)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
1..1096
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="008X04"
/clone_lib="A"
/notes="Genoscope sequence ID : C0AA008BF02c1-end : T7"

BASE COUNT 208 a 286 c 361 g 226 t 15 others

ORIGIN

alignment_scores:
Quality: 217.00 Length: 109
Ratio: 2.855 Gaps: 2
Percent Similarity: 69.725 Percent Identity: 44.037

alignment_block:
US-09-775-181-4 x CNS05CMB/rev ..

Align seg 1/1 to reverse of: CNS05CMB from: 1 to: 1096

194 AlAtHrArgGluUSeSerArgIletLeuLeuGlnAsPleuSerSerAl 210
:::||||| ::::::::::::::::::::|||

1037 GCGACCGGGCGGCAGCATCTCTCGACGACCTTCGACGGTGCC 988
:::||||| ::::::::::::::::::::|||

210 aProHisLeuAlaAsnLaThrLeuGluThrGluTrpPheHisGlyLeuA 227
:::||||| ||| ||| ::::||| ::::||| ;

987 CGGGGGGTCTCACACCACCGCCGCCGCGGTNGTACCACGACGCCA 938
:::||||| ::::::::::::::::::::|||

227 rgArg...LyTrpArgProHisLeuHisArgArgGlyProAsnGlnGly 242
:::||||| ::::::::::::::::::::|||

937 AGGAGCGCGAGAAGAGCCCGCTTCGGAAGAGGGTCTCGACCGACAAC 888
:::||||| ::::::::::::::::::::|||

243 ProArgGlyLeuGlyHisSerTrpArgArgLysaspGlyLeuGlyGlyAs 259
||| ::::||| ||| ::::||| ::::||| |||

887 CCG...AGCCTAGATCTGTGGCGGGGAGCCAGCTTTGCCCGGA 841
:::||||| ::::::::::::::::::::|||

259 pLYSerHisPhelysTrpSerProTyrrLeuGluCysGluAsnGlys 276
1:::||||| ::::::::::::::::::::|||

840 CAGGCGCGACGTAGCTGTCTGGCGCCCTACTGTGAGTGTGGGAGGCG 791
:::||||| ::::::::::::::::::::|||

276 eTrYrLYsProGlyTrpLeuValThrLeuSerSerAlaIetyrclyLeu 292
||| ::::::::::::::::::::|||

790 TTCGTCTCCCAGCTGGCTCTGACCTTATCGGCTGCCCCCTCTACGGCCTG 741
:::||||| ::::::::::::::::::::|||

293 GlnProAsnLeuValProGluPheArg 301
:::||||| ::::::::::::::::::::|||

740 AAGACGACCTGGCTCCGGAGTTCAGG 714
:::||||| ::::::::::::::::::::|||


```

VERSION      A2628814.1  GI:11751004
KEYWORDS     GSS.
SOURCE       house mouse.
ORGANISM     Mus musculus
REFERENCE    1 (bases 1 to 329)
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
             and Wright,D., Weiss,R.
TITLE        Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
             University of Utah Genome Center
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000 Std Error: 0.00
             Plate: 0481 row: D column: 17
             Seq primer: GGTGTAAACGACGCGCAGT
             Class: plasmid ends
             High quality sequence stop: 329.
FEATURES     source
             1..329
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0481D17"
                /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
             Laboratory Mouse DNA Resource
             (http://www.jax.org/resources/documents/dnares/). The DNA
             was hydrodynamically sheared by repeated passage through a
             0.005 inch orifice at constant velocity. The sheared DNA
             was blunt end-repaired with T4 DNA polymerase and T4
             polynucleotide kinase. Adaptor oligonucleotides were
             ligated to the blunt ends in high molar excess. The
             adaptor DNA was purified and size-selected for a 9.5 to
             10.5 kb range using preparative agarose gel
             electrophoresis. Vector DNA was prepared from a derivative
             of PWD42 (gil4732114[gblAF129072.1]), a copy-number
             inducible derivative of plasmid R1. The vector was ligated
             with adaptors complementary to the insert adaptors and
             purified. The sheared, adaptor mouse DNA was annealed to
             adaptor vector DNA, and transformed into
             chemically-competent E. coli XL10-Gold (Stratagene) cells
             and selected for ampicillin resistance."
BASE COUNT   98 a 55 c 80 g 96 t
ORIGIN
alignment_scores:
    Quality: 201.00      Length: 37
    Ratio: 5.432         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.297
alignment_block:
US-09-775-181-4 x A2628814
Align seg 1/1 to: A2628814 from: 1 to: 329
300 PheArgGlyValMetLysValAspLeuAsnLeuGlnLysValAspLeuAs 316
|||||
40 TTTAGGGGTGTGTAAGCTTGTATTAACCTTCAGAAAGTGGACATTGA 89

```

```

316 pGlnCysSerSerAspGlyTrpPheSerGlyThrHisLysCysHisLeuA 333
|||||
90 CCAATGTTTCGAGTATGGCTGTTTTCAGGAAGTCCAAATGCCACCTTA 139
333 snAsnSerGlu 336
|||||
140 ACAACTCAGAG 150
seq_name: gb_est27:AJ280627
seq_documentation_block:
LOCUS      AJ280627      518 bp      mRNA      EST      30-JUN-2000
DEFINITION 4A3A-AAR-F-11-F Anopheles gambiae immune competent 4A3A Anopheles
gambiae cDNA clone 4A3A-AAR-F-11, mRNA sequence.
ACCESSION  AJ280627
VERSION     AJ280627.1 GI:6928508
KEYWORDS    EST.
SOURCE      African malaria mosquito.
ORGANISM    Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
            ; Anopheles.
REFERENCE   1 (bases 1 to 518)
AUTHORS     Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
            Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B.
            and Kafatos,F.C.
TITLE        Anopheles gambiae pilot gene discovery project: identification of
            mosquito innate immunity genes from expressed sequence tags
            generated from immune-competent cell lines
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)
MEDLINE     20300950
COMMENT     Contact: Dimopoulos G
            Fotis C. Kafatos laboratory
            European Molecular Biology Laboratory
            Meyerhofstrasse 1, 69117 Heidelberg, Germany.
FEATURES     source
            1..518
                /organism="Anopheles gambiae"
                /strain="4A t/r"
                /db_xref="taxon:7165"
                /clone="4A3A-AAR-F-11"
                /clone_lib="Anopheles gambiae immune competent 4A3A"
                /cell_line="Immune competent 4A3A"
                /lab_host="E. coli DH10B"
                /note="vector: p773D-Pac (Pharmacia) with a modified
            polylinker; Site 1: EcoRI; Site 2: NotI; sequenced from
            forward priming site which reads from the 3' end of the
            cDNA. The 4A3A is a directionally cloned and normalized
            cDNA library that was constructed from the 4A3A cell line
            oligo-T primed cDNA according to: Bonaldo, Lennon & Soares
            (1996) : Normalization and Subtraction: Two approaches To
            Facilitate Gene Discovery, Genome Research 6, 791-806."
BASE COUNT   109 a 149 c 129 g 130 t 1 others
ORIGIN
alignment_scores:
    Quality: 200.00      Length: 154
    Ratio: 2.299         Gaps: 6
Percent Similarity: 56.494 Percent Identity: 31.169
alignment_block:
US-09-775-181-4 x AJ280627/rev ..
Align seg 1/1 to reverse of: AJ280627 from: 1 to: 518
262 HisPheLysTrpSerProTyrLeuGluCysGluAsnGlySerTyrly 278
|||||
420 CACGGCTACTGGACGACGCCGCGAGTTCGACTGCAAG.....GGCTACGT 377
278 sProGlyTrpLeuValThrLeuSerSerAlaIleTyrGly.....L 292
|||||
376 GAAGAAGTGGCTCATAAACGTACGGGTGCCGTTCTTCGGCTGGACAGCC 327

```

```
292 euGlnProAsnLeuValProGluPheArgGlyValMetLysValAspIle 308
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
326 TCAAGCGCAGCTA.....GAATTCAAGGAGTCGTTGCAGTTTCAATG 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 AsnLeuGlnLysValAspIleAspGlnCysSerSerAspGlyTyr..... 323
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 AACATGCTCCAGCTGGGATATCAACCAATGTCGGGATGATTATTACGAGCC 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 .....PheSerGlyThrHisLysCysHisLeuAsnAsnSerGluCysM 338
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 GAATGCATTCAAAACACTCACAATCGGACGAGAGCTTTCATACTGTG 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
338 etProIleLysGlyLeuGlyPheValLeuGlyAlaTyrGluCysIleCys 354
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 TCCCGATCTGGCGCGGGTACGAAACGGCGGCTATAAGTGGAGTGC 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 LysAlaGlyPhe.....TyrHisProGlu 362
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 CTGAAGGGTTCGAGTATCCGTACGAGATCTGATCACCTACTATGACGG 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 yValLeuProValAsnAsnPheArgArgGlyProAspGlnHisIleS 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 GCAGCTGGTCGAGTCCGAGTTCGAGACATCGTGATGATAAG..... 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 erGlySerThrLysAspValSerGluGluAlaTyrValCysLeuProCys 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 .....CAATCGCGGTTCGAGACGTTCAAGTGC 13
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 ArgGluGlyCys 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
12 CGTCTGGCTTGT 1
```